



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>7</sup> :</b>  <b>C07K 14/00</b>	<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 00/24765</b>  <b>(43) International Publication Date:</b> 4 May 2000 (04.05.00)																																							
<div style="display: flex; justify-content: space-between;"> <div style="width: 48%;"> <b>(21) International Application Number:</b> PCT/CA99/00992   <b>(22) International Filing Date:</b> 28 October 1999 (28.10.99)   <b>(30) Priority Data:</b> <table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 30%;">60/106,034</td><td style="width: 40%;">28 October 1998 (28.10.98)</td><td style="width: 30%;">US</td></tr> <tr><td>60/106,044</td><td>28 October 1998 (28.10.98)</td><td>US</td></tr> <tr><td>60/106,039</td><td>28 October 1998 (28.10.98)</td><td>US</td></tr> <tr><td>60/106,042</td><td>28 October 1998 (28.10.98)</td><td>US</td></tr> <tr><td>60/106,087</td><td>29 October 1998 (29.10.98)</td><td>US</td></tr> <tr><td>60/106,072</td><td>29 October 1998 (29.10.98)</td><td>US</td></tr> <tr><td>60/106,073</td><td>29 October 1998 (29.10.98)</td><td>US</td></tr> <tr><td>60/106,074</td><td>29 October 1998 (29.10.98)</td><td>US</td></tr> <tr><td>60/106,589</td><td>2 November 1998 (02.11.98)</td><td>US</td></tr> <tr><td>60/107,034</td><td>2 November 1998 (02.11.98)</td><td>US</td></tr> <tr><td>60/107,035</td><td>2 November 1998 (02.11.98)</td><td>US</td></tr> <tr><td>60/106,587</td><td>2 November 1998 (02.11.98)</td><td>US</td></tr> <tr><td>60/106,588</td><td>2 November 1998 (02.11.98)</td><td>US</td></tr> </table> </div> <div style="width: 48%;"> <b>(74) Agents:</b> MORROW, Joy, D. et al.; Smart &amp; Biggar, 900-55 Metcalfe Street, P.O. Box 2999, Station D, Ottawa, Ontario K1P 5Y6 (CA).   <b>(81) Designated States:</b> AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).         </div> </div>			60/106,034	28 October 1998 (28.10.98)	US	60/106,044	28 October 1998 (28.10.98)	US	60/106,039	28 October 1998 (28.10.98)	US	60/106,042	28 October 1998 (28.10.98)	US	60/106,087	29 October 1998 (29.10.98)	US	60/106,072	29 October 1998 (29.10.98)	US	60/106,073	29 October 1998 (29.10.98)	US	60/106,074	29 October 1998 (29.10.98)	US	60/106,589	2 November 1998 (02.11.98)	US	60/107,034	2 November 1998 (02.11.98)	US	60/107,035	2 November 1998 (02.11.98)	US	60/106,587	2 November 1998 (02.11.98)	US	60/106,588	2 November 1998 (02.11.98)	US
60/106,034	28 October 1998 (28.10.98)	US																																							
60/106,044	28 October 1998 (28.10.98)	US																																							
60/106,039	28 October 1998 (28.10.98)	US																																							
60/106,042	28 October 1998 (28.10.98)	US																																							
60/106,087	29 October 1998 (29.10.98)	US																																							
60/106,072	29 October 1998 (29.10.98)	US																																							
60/106,073	29 October 1998 (29.10.98)	US																																							
60/106,074	29 October 1998 (29.10.98)	US																																							
60/106,589	2 November 1998 (02.11.98)	US																																							
60/107,034	2 November 1998 (02.11.98)	US																																							
60/107,035	2 November 1998 (02.11.98)	US																																							
60/106,587	2 November 1998 (02.11.98)	US																																							
60/106,588	2 November 1998 (02.11.98)	US																																							
<div style="display: flex; justify-content: space-between;"> <div style="width: 48%;"> <b>(71) Applicant (for all designated States except US):</b> CONNAUGHT LABORATORIES LIMITED [CA/CA]; 1755 Steeles Avenue West, Toronto, Ontario M2R 3T4 (CA).   <b>(72) Inventors; and</b>  <b>(75) Inventors/Applicants (for US only):</b> MURDIN, Andrew, D. [CA/CA]; 146 Rhodes Circle, Newmarket, Ontario L3X         </div> <div style="width: 48%;"> <b>Published</b>  <i>Without international search report and to be republished upon receipt of that report.</i> </div> </div>																																									
<b>(54) Title:</b> CHLAMYDIA ANTIGENES AND CORRESPONDING DNA FRAGMENTS AND USES THEREOF																																									
<b>(57) Abstract</b>  <p>The present invention provides purified and isolated polynucleotide molecules that encode <i>Chlamydia</i> polypeptides which can be used in methods to prevent, treat, and diagnose <i>Chlamydia</i> infection. In one form of the invention, the polynucleotide molecules are selected from DNA that encode polypeptides CPN100397 (SEQ ID Nos: 1 and 2), CPN100421 (SEQ ID Nos: 3 and 4), CPN100422 (SEQ ID Nos: 4 and 6), CPN100424 (SEQ ID Nos: 7 and 8), CPN100426 (SEQ ID Nos: 9 and 10), CPN100508 (SEQ ID Nos: 11 and 12), CPN100515 (SEQ ID Nos: 13 and 14), CPN100538 (SEQ ID Nos: 15 and 16), CPN100557 (SEQ ID Nos: 17 and 18), CPN100622 (SEQ ID Nos: 19 and 20), CPN100626 (SEQ ID Nos: 21 and 22), CPN100628 (SEQ ID Nos: 23 and 24) and CPN100630 (SEQ ID Nos: 25 and 26).</p>																																									

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

**TITLE OF INVENTION**

CHLAMYDIA ANTIGENS AND CORRESPONDING DNA FRAGMENTS AND USES  
THEREOF

**5 REFERENCE TO RELATED APPLICATIONS**

This application claims the benefit of U.S.  
Provisional Application No. 60/106034, filed October 28, 1998,  
U.S. Provisional Application No. 60/106039, filed October 28,  
1998, U.S. Provisional Application No. 60/106042, filed October  
10 28, 1998, U.S. Provisional Application No. 60/106044, filed  
October 28, 1998, U.S. Provisional Application No. 60/106072,  
filed October 29, 1998, U.S. Provisional Application No.  
60/106073, filed October 29, 1998, U.S. Provisional Application  
No. 60/106074, filed October 29, 1998, U.S. Provisional  
15 Application No. 60/106087, filed October 29, 1998, U.S.  
Provisional Application No. 60/106587, filed November 2, 1998,  
U.S. Provisional Application No. 60/106588, filed November 2,  
1998, U.S. Provisional Application No. 60/107089, filed November  
2, 1998, U.S. Provisional Application No. 60/107034, filed  
20 November 2, 1998 and U.S. Provisional Application No. 60/107035,  
filed November 2, 1998.

**FIELD OF INVENTION**

The present invention relates to *Chlamydia* antigens  
25 and corresponding DNA molecules, which can be used to prevent  
and treat *Chlamydia* infection in mammals, such as humans.

**BACKGROUND OF THE INVENTION**

Chlamydiae are prokaryotes. They exhibit morphologic  
30 and structural similarities to gram-negative bacteria including  
a trilaminar outer membrane, which contains lipopolysaccharide  
and several membrane proteins that are structurally and  
functionally analogous to proteins found in *E coli*. They are  
obligate intra-cellular parasites with a unique biphasic life

cycle consisting of a metabolically inactive but infectious extracellular stage and a replicating but non-infectious intracellular stage. The replicative stage of the life-cycle takes place within a membrane-bound inclusion which sequesters the bacteria away from the cytoplasm of the infected host cell.

*C. pneumoniae* is a common human pathogen, originally described as the TWAR strain of *Chlamydia psittaci* but subsequently recognised to be a new species. *C. pneumoniae* is antigenically, genetically and morphologically distinct from other chlamydia species (*C. trachomatis*, *C. pecorum* and *C. psittaci*). It shows 10% or less DNA sequence homology with either of *C. trachomatis* or *C. psittaci*.

*C. pneumoniae* is a common cause of community acquired pneumonia, only less frequent than *Streptococcus pneumoniae* and *Mycoplasma pneumoniae* (Grayston et al. (1995) Journal of Infectious Diseases 168:1231; Campos et al. (1995) Investigation of Ophthalmology and Visual Science 36:1477). It can also cause upper respiratory tract symptoms and disease, including bronchitis and sinusitis (Grayston et al. (1995) Journal of Infectious Diseases 168:1231; Grayston et al (1990) Journal of Infectious Diseases 161:618; Marrie (1993) Clinical Infectious Diseases. 18:501; Wang et al (1986) Chlamydial infections). Cambridge University Press, Cambridge. p. 329 The great majority of the adult population (over 60%) has antibodies to *C. pneumoniae* (Wang et al (1986) Chlamydial infections. Cambridge University Press, Cambridge. p. 329), indicating past infection which was unrecognized or asymptomatic.

*C. pneumoniae* infection usually presents as an acute respiratory disease (i.e., cough, sore throat, hoarseness, and fever; abnormal chest sounds on auscultation). For most patients, the cough persists for 2 to 6 weeks, and recovery is slow. In approximately 10% of these cases, upper respiratory tract infection is followed by bronchitis or pneumonia. Furthermore, during a *C. pneumoniae* epidemic, subsequent

co-infection with pneumococcus has been noted in about half of these pneumonia patients, particularly in the infirm and the elderly. As noted above, there is more and more evidence that *C. pneumoniae* infection is also linked to diseases other than  
5 respiratory infections.

The reservoir for the organism is presumably people. In contrast to *C. psittaci* infections, there is no known bird or animal reservoir. Transmission has not been clearly defined. It may result from direct contact with secretions, from fomites, or  
10 from airborne spread. There is a long incubation period, which may last for many months. Based on analysis of epidemics, *C. pneumoniae* appears to spread slowly through a population (case-to-case interval averaging 30 days) because infected persons are inefficient transmitters of the organism. Susceptibility to *C.*  
15 *pneumoniae* is universal. Reinfections occur during adulthood, following the primary infection as a child. *C. pneumoniae* appears to be an endemic disease throughout the world, noteworthy for superimposed intervals of increased incidence (epidemics) that persist for 2 to 3 years. *C. trachomatis*  
20 infection does not confer cross-immunity to *C. pneumoniae*. Infections are easily treated with oral antibiotics, tetracycline or erythromycin (2 g/d, for at least 10 to 14 d). A recently developed drug, azithromycin, is highly effective as a single-dose therapy against chlamydial infections.

25 In most instances, *C. pneumoniae* infection is often mild and without complications, and up to 90% of infections are subacute or unrecognized. Among children in industrialized countries, infections have been thought to be rare up to the age of 5 y, although a recent study (E Normann et al, Chlamydia  
30 *pneumoniae* in children with acute respiratory tract infections, Acta Paediatrica, 1998, Vol 87, Iss 1, pp 23-27) has reported that many children in this age group show PCR evidence of infection despite being seronegative, and estimates a prevalence of 17-19% in 2-4 y olds. In developing countries, the

seroprevalence of *C. pneumoniae* antibodies among young children is elevated, and there are suspicions that *C. pneumoniae* may be an important cause of acute lower respiratory tract disease and mortality for infants and children in tropical regions of the world.

From seroprevalence studies and studies of local epidemics, the initial *C. pneumoniae* infection usually happens between the ages of 5 and 20 y. In the USA, for example, there are estimated to be 30,000 cases of childhood pneumonia each year caused by *C. pneumoniae*. Infections may cluster among groups of children or young adults (e.g., school pupils or military conscripts).

*C. pneumoniae* causes 10 to 25% of community-acquired lower respiratory tract infections (as reported from Sweden, Italy, Finland, and the USA). During an epidemic, *C. pneumoniae* infection may account for 50 to 60% of the cases of pneumonia. During these periods, also, more episodes of mixed infections with *S. pneumoniae* have been reported.

Reinfection during adulthood is common; the clinical presentation tends to be milder. Based on population seroprevalence studies, there tends to be increased exposure with age, which is particularly evident among men. Some investigators have speculated that a persistent, asymptomatic *C. pneumoniae* infection state is common.

In adults of middle age or older, *C. pneumoniae* infection may progress to chronic bronchitis and sinusitis. A study in the USA revealed that the incidence of pneumonia caused by *C. pneumoniae* in persons younger than 60 years is 1 case per 1,000 persons per year; but in the elderly, the disease incidence rose three-fold. *C. pneumoniae* infection rarely leads to hospitalization, except in patients with an underlying illness.

Of considerable importance is the association of atherosclerosis and *C. pneumoniae* infection. There are several

epidemiological studies showing a correlation of previous infections with *C. pneumoniae* and heart attacks, coronary artery and carotid artery disease (Saikku et al. (1988) Lancet;ii:983; Thom et al. (1992) JAMA 268:68; Linnanmaki et al. (1993),  
5 Circulation 87:1030; Saikku et al. (1992) Annals Internal Medicine 116:273; Melnick et al (1993) American Journal of Medicine 95:499). Moreover, the organisms has been detected in atheromas and fatty streaks of the coronary, carotid, peripheral arteries and aorta (Shor et al. (1992) South African. Medical  
10 Journal 82:158; Kuo et al. (1993) Journal of Infectious Diseases 167:841; Kuo et al. (1993) Arteriosclerosis and Thrombosis 13:1500; Campbell et al (1995) Journal of Infectious Diseases 172:585; Chiu et al. Circulation, 1997 (In Press)). Viable *C. pneumoniae* has been recovered from the coronary and carotid  
15 artery (Ramirez et al (1996) Annals of Internal Medicine 125:979; Jackson et al. Abst. K121, p272, 36<sup>th</sup> ICAAC, 15-18 Sept. 1996, New Orleans). Furthermore, it has been shown that *C. pneumoniae* can induce changes of atherosclerosis in a rabbit model (Fong et al (1997) Journal of Clinical Microbiology  
20 35:48). Taken together, these results indicate that it is highly probable that *C. pneumoniae* can cause atherosclerosis in humans, though the epidemiological importance of chlamydial atherosclerosis remains to be demonstrated.

A number of recent studies have also indicated an  
25 association between *C. pneumoniae* infection and asthma. Infection has been linked to wheezing, asthmatic bronchitis, adult-onset asthma and acute exacerbations of asthma in adults, and small-scale studies have shown that prolonged antibiotic treatment was effective at greatly reducing the severity of the  
30 disease in some individuals (Hahn DL, et al. Evidence for Chlamydia pneumoniae infection in steroid-dependent asthma. Ann Allergy Asthma Immunol. 1998 Jan; 80(1): 45-49.; Hahn DL, et al. Association of Chlamydia pneumoniae IgA antibodies with recently symptomatic asthma. Epidemiol Infect. 1996 Dec;

117(3): 513-517; Bjornsson E, et al. Serology of chlamydia in relation to asthma and bronchial hyperresponsiveness. Scand J Infect Dis. 1996; 28(1): 63-69.; Hahn DL. Treatment of *Chlamydia pneumoniae* infection in adult asthma: a before-after trial. J Fam Pract. 1995 Oct; 41(4): 345-351.; Allegra L, et al. Acute exacerbations of asthma in adults: role of *Chlamydia pneumoniae* infection. Eur Respir J. 1994 Dec; 7(12): 2165-2168.; Hahn DL, et al. Association of *Chlamydia pneumoniae* (strain TWAR) infection with wheezing, asthmatic bronchitis, and adult-onset asthma. JAMA. 1991 Jul 10; 266(2): 225-230).

In light of these results a protective vaccine against *C. pneumoniae* infection would be of considerable importance. There is not yet an effective vaccine for any human chlamydial infection. It is conceivable that an effective vaccine can be developed using physically or chemically inactivated *Chlamydiae*. However, such a vaccine does not have a high margin of safety. In general, safer vaccines are made by genetically manipulating the organism by attenuation or by recombinant means. Accordingly, a major obstacle in creating an effective and safe vaccine against human chlamydial infection has been the paucity of genetic information regarding *Chlamydia*, specifically *C. pneumoniae*.

Studies with *C. trachomatis* and *C. psittaci* indicate that safe and effective vaccine against *Chlamydia* is an attainable goal. For example, mice which have recovered from a lung infection with *C. trachomatis* are protected from infertility induced by a subsequent vaginal challenge (Pal et al. (1996) Infection and Immunity. 64:5341). Similarly, sheep immunized with inactivated *C. psittaci* were protected from subsequent chlamydial-induced abortions and stillbirths (Jones et al. (1995) Vaccine 13:715). Protection from chlamydial infections has been associated with Th1 immune responses, particularly the induction of INF $\gamma$  - producing CD4+T-cells (Igiertemes et al. (1993) Immunology 5:317). The adoptive



transfer of CD4+ cell lines or clones to nude or SCID mice conferred protection from challenge or cleared chronic disease (Igietseme et al (1993) Regional Immunology 5:317; Magee et al (1993) Regional Immunology 5: 305), and in vivo depletion of CD4+ T cells exacerbated disease post-challenge (Landers et al (1991) Infection & Immunity 59:3774; Magee et al (1995) Infection & Immunity 63:516). However, the presence of sufficiently high titres of neutralising antibody at mucosal surfaces can also exert a protective effect (Cotter et al. (1995) Infection and Immunity 63:4704).

Antigenic variation within the species *C. pneumoniae* is not well documented due to insufficient genetic information, though variation is expected to exist based on *C. trachomatis*. Serovars of *C. trachomatis* are defined on the basis of antigenic variation in MOMP, but published *C. pneumoniae* MOMP gene sequences show no variation between several diverse isolates of the organism (Campbell et al (1990) Infection and Immunity 58:93; McCafferty et al (1995) Infection and Immunity 63:2387-9; Knudsen et al (1996) Third Meeting of the European Society for Chlamydia Research, Vienna). Regions of the protein known to be conserved in other chlamydial MOMPs are conserved in *C. pneumoniae* (Campbell et al (1990) Infection and Immunity 58:93; McCafferty et al (1995) Infection and Immunity 63:2387-9). One study has described a strain of *C. pneumoniae* with a MOMP of greater than usual molecular weight, but the gene for this has not been sequenced (Grayston et al. (1995) Journal of Infectious Diseases 168:1231). Partial sequences of outer membrane protein 2 from nine diverse isolates were also found to be invariant (Ramirez et al (1996) Annals of Internal Medicine 125:979). The genes for HSP60 and HSP70 show little variation from other chlamydial species, as would be expected. The gene encoding a 76kDa antigen has been cloned from a single strain of *C. pneumoniae*. It has no significant similarity with other known

chlamydial genes (Marrie (1993) Clinical Infectious Diseases. 18:501).

Many antigens recognised by immune sera to *C. pneumoniae* are conserved across all chlamydiae, but 98kDa, 76 kDa and 54 kDa proteins appear to be *C. pneumoniae*-specific (Ref Campos et al. (1995) Investigation of Ophthalmology and Visual Science 36:1477; Marrie (1993) Clinical Infectious Diseases. 18:501; Wiedmann-Al-Ahmad M, et al. Reactions of polyclonal and neutralizing anti-p54 monoclonal antibodies with an isolated, species-specific 54-kilodalton protein of *Chlamydia pneumoniae*. Clin Diagn Lab Immunol. 1997 Nov; 4(6): 700-704). Immunoblotting of isolates with sera from patients does show variation of blotting patterns between isolates, indicating that serotypes *C. pneumoniae* may exist (Ref 1,16). However, the results are potentially confounded by the infection status of the patients, since immunoblot profiles of a patient's sera change with time post-infection. An assessment of the number and relative frequency of any serotypes, and the defining antigens, is not yet possible.

Accordingly, a need exists for identifying and isolating polynucleotide sequences of *C. pneumoniae* for use in preventing and treating *Chlamydia* infection.

#### **SUMMARY OF THE INVENTION**

The present invention provides purified and isolated polynucleotide molecules that encode *Chlamydia* polypeptides which can be used in methods to prevent, treat, and diagnose *Chlamydia* infection. In one form of the invention, the polynucleotide molecules are selected from DNA that encode polypeptides CPN100397 (SEQ ID Nos: 1 and 2), CPN100421 (SEQ ID Nos: 3 and 4), CPN100422 (SEQ ID Nos: 5 and 6), CPN100424 (SEQ ID Nos: 7 and 8), CPN100426 (SEQ ID Nos: 9 and 10), CPN100508 (SEQ ID Nos: 11 and 12), CPN100515 (SEQ ID Nos: 13 and 14), CPN100538 (SEQ ID Nos: 15 and 16), CPN100557 (SEQ ID Nos: 17 and

18), CPN100622 (SEQ ID Nos: 19 and 20), CPN100626 (SEQ ID Nos: 21 and, 22), CPN100628 (SEQ ID Nos: 23 and 24) and CPN100630 (SEQ ID Nos: 25 and 26).

Another form of the invention provides polypeptides  
5 corresponding to the isolated DNA molecules. The amino acid sequences of the corresponding encoded polypeptides are shown for CPN100397 as SEQ ID Nos: 27 and 28, CPN100421 as SEQ ID No: 29, CPN100422 as SEQ ID No: 30, CPN100424 as SEQ ID No: 31, CPN100426 as SEQ ID No: 32, CPN100508 as SEQ ID Nos: 33 and 34,  
10 CPN100515 as SEQ ID Nos: 35 and 36, CPN100538 as SEQ ID No: 37, CPN100557 as SEQ ID Nos: 38 and 39, CPN100622 as SEQ ID Nos: 40 and 41, CPN100626 as SEQ ID No: 42, CPN100628 as SEQ ID No: 43 and CPN100630 as SEQ ID Nos: 44 and 45.

Those skilled in the art will readily understand that the  
15 invention, having provided the polynucleotide sequences encoding *Chlamydia* polypeptides, also provides polynucleotides encoding fragments derived from such peptides. Moreover, the invention is understood to provide mutants and derivatives of such polypeptides and fragments derived therefrom, which result from  
20 the addition, deletion, or substitution of non-essential amino acids as described herein. Those skilled in the art would also readily understand that the invention, having provided the polynucleotide sequences encoding *Chlamydia* polypeptides, further provides monospecific antibodies that specifically bind  
25 to such polypeptides

The present invention has wide application and includes expression cassettes, vectors, and cells transformed or transfected with the polynucleotides of the invention. Accordingly, the present invention further provides (i) a method  
30 for producing a polypeptide of the invention in a recombinant host system and related expression cassettes, vectors, and transformed or transfected cells; (ii) a vaccine, or a live vaccine vector such as a pox virus, *Salmonella typhimurium*, or *Vibrio cholerae* vector, containing a polynucleotide of the

invention, such vaccines and vaccine vectors being useful for, e.g., preventing and treating *Chlamydia* infection, in combination with a diluent or carrier, and related pharmaceutical compositions and associated therapeutic and/or prophylactic methods; (iii) a therapeutic and/or prophylactic use of an RNA or DNA molecule of the invention, either in a naked form or formulated with a delivery vehicle, a polypeptide or combination of polypeptides, or a monospecific antibody of the invention, and related pharmaceutical compositions; (iv) a method for diagnosing the presence of *Chlamydia* in a biological sample, which can involve the use of a DNA or RNA molecule, a monospecific antibody, or a polypeptide of the invention; and (v) a method for purifying a polypeptide of the invention by antibody-based affinity chromatography.

15

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

The present invention will be further understood from the following description with reference to the drawings, in which:

Figure 1 shows the nucleotide sequence of the CPN100397 (SEQ ID No: 1 - entire sequence and SEQ ID No: 2 - coding sequence) and the deduced amino acid sequence of the CPN100397 protein from *Chlamydia pneumoniae* (SEQ ID No: 27 and 28).

Figure 2 shows the restriction enzyme analysis of the gene encoding the *C. pneumoniae* CPN100397 gene.

Figure 3 shows the nucleotide sequence of the CPN100421 (SEQ ID No: 3 - entire sequence and SEQ ID No: 4 - coding sequence) and the deduced amino acid sequence of the CPN100421 protein from *Chlamydia pneumoniae* (SEQ ID No: 29).

Figure 4 shows the restriction enzyme analysis of the gene encoding the *C. pneumoniae* CPN100421 gene.

Figure 5 shows the nucleotide sequence of the CPN100422 (SEQ ID No: 5 - entire sequence and SEQ ID No: 6 - coding sequence) and the deduced amino acid sequence of the CPN100422 protein from *Chlamydia pneumoniae* (SEQ ID No: 30).

Figure 6 shows the restriction enzyme analysis of the gene encoding the *C. pneumoniae* CPN100422 gene.

Figure 7 shows the nucleotide sequence of the CPN100424 (SEQ ID No: 7 - entire sequence and SEQ ID No: 8 - coding sequence) and the deduced amino acid sequence of the CPN100424 protein from *Chlamydia pneumoniae* (SEQ ID No: 31).

Figure 8 shows the restriction enzyme analysis of the gene encoding the *C. pneumoniae* CPN100424 gene.

Figure 9 shows the nucleotide sequence of the CPN100426 (SEQ ID No: 9 - entire sequence and SEQ ID No: 10 - coding sequence) and the deduced amino acid sequence of the CPN100426 protein from *Chlamydia pneumoniae* (SEQ ID No: 32).

Figure 10 shows the restriction enzyme analysis of the gene encoding the *C. pneumoniae* CPN100426 gene.

Figure 11 shows the nucleotide sequence of the CPN100508 (SEQ ID No: 11 - entire sequence and SEQ ID No: 12 - coding sequence) and the deduced amino acid sequence of the CPN100508 protein from *Chlamydia pneumoniae* (SEQ ID No: 33 - full length sequence and SEQ ID No: 34 - processed sequence).

Figure 12 shows the restriction enzyme analysis of the gene encoding the *C. pneumoniae* CPN100508 gene.

Figure 13 shows the nucleotide sequence of the CPN100515 (SEQ ID No: 13 - entire sequence and SEQ ID No: 14 - coding sequence) and the deduced amino acid sequence of the CPN100515 protein from *Chlamydia pneumoniae* (SEQ ID No: 35 - full length sequence and SEQ ID No: 36 - processed sequence).

Figure 14 shows the restriction enzyme analysis of the gene encoding the *C. pneumoniae* CPN100515 gene.

Figure 15 shows the nucleotide sequence of the CPN100538 (SEQ ID No: 15 - entire sequence and SEQ ID No: 16 - coding sequence) and the deduced amino acid sequence of the CPN100538 protein from *Chlamydia pneumoniae* (SEQ ID No: 37).

Figure 16 shows the restriction enzyme analysis of the gene encoding the *C. pneumoniae* CPN100538 gene.

Figure 17 shows the nucleotide sequence of the CPN100557 (SEQ ID No: 17 - entire sequence and SEQ ID No: 18 - coding sequence) and the deduced amino acid sequence of the CPN100557 protein from *Chlamydia pneumoniae* (SEQ ID No: 38 - full length 5 sequence and SEQ ID No: 39 - processed sequence).

Figure 18 shows the restriction enzyme analysis of the gene encoding the *C. pneumoniae* CPN100557 gene.

Figure 19 shows the nucleotide sequence of the CPN100622 (SEQ ID No: 19 - entire sequence and SEQ ID No: 20 - coding 10 sequence) and the deduced amino acid sequence of the CPN100622 protein from *Chlamydia pneumoniae* (SEQ ID No: 40 - full length sequence and SEQ ID No: 41 - processed sequence).

Figure 20 shows the restriction enzyme analysis of the gene encoding the *C. pneumoniae* CPN100622 gene.

15 Figure 21 shows the nucleotide sequence of the CPN100626 (SEQ ID No: 21 - entire sequence and SEQ ID No: 22 - coding sequence) and the deduced amino acid sequence of the CPN100626 protein from *Chlamydia pneumoniae* (SEQ ID No: 42).

Figure 22 shows the restriction enzyme analysis of the 20 gene encoding the *C. pneumoniae* CPN100626 gene.

Figure 23 shows the nucleotide sequence of the CPN100628 (SEQ ID No: 23 - entire sequence and SEQ ID No: 24 - coding sequence) and the deduced amino acid sequence of the CPN100628 protein from *Chlamydia pneumoniae* (SEQ ID No: 43).

25 Figure 24 shows the restriction enzyme analysis of the gene encoding the *C. pneumoniae* CPN100628 gene.

Figure 25 shows the nucleotide sequence of the CPN100630 (SEQ ID No: 25 - entire sequence and SEQ ID No: 26 - coding sequence) and the deduced amino acid sequence of the CPN100630 30 protein from *Chlamydia pneumoniae* (SEQ ID No: 44 - full length sequence and SEQ ID No: 45 - processed sequence).

Figure 26 shows the restriction enzyme analysis of the gene encoding the *C. pneumoniae* CPN100630 gene.

Figures 27 through 39 show an identification of T and B cell epitopes from the amino acid sequences shown in the foregoing figures.

## 5 DETAILED DESCRIPTION OF INVENTION

Open reading frames (ORFs) encoding chlamydial polypeptides have been identified from the *C. pneumoniae* genome. These polypeptides include polypeptides found permanently in the bacterial membrane structure, polypeptides present in the  
10 external vicinity of the bacterial membrane, polypeptides found permanently in the inclusion membrane structure, polypeptides present in the external vicinity of the inclusion membrane, and polypeptides released into the cytoplasm of the infected cell. These polypeptides can be used to prevent and treat *Chlamydia*  
15 infection.

According to a first aspect of the invention, isolated polynucleotides are provided which encode the precursor and mature forms of *Chlamydia* polypeptides, whose amino acid sequences are selected from the group consisting of: SEQ ID  
20 Nos: 27 to 45.

The term "isolated polynucleotide" is defined as a polynucleotide removed from the environment in which it naturally occurs. For example, a naturally-occurring DNA molecule present in the genome of a living bacteria or as part  
25 of a gene bank is not isolated, but the same molecule separated from the remaining part of the bacterial genome, as a result of, e.g., a cloning event (amplification), is isolated. Typically, an isolated DNA molecule is free from DNA regions (e.g., coding regions) with which it is immediately contiguous at the 5' or 3'  
30 end, in the naturally occurring genome. Such isolated polynucleotides may be part of a vector or a composition and still be defined as isolated in that such a vector or composition is not part of the natural environment of such polynucleotide.

The polynucleotide of the invention is either RNA or DNA (cDNA, genomic DNA, or synthetic DNA), or modifications, variants, homologs or fragments thereof. The DNA is either double-stranded or single-stranded, and, if single-stranded, is either the coding strand or the non-coding (anti-sense) strand. Any one of the sequences that encode the polypeptides of the invention as shown in SEQ ID Nos: 1 to 26 is (a) a coding sequence, (b) a ribonucleotide sequence derived from transcription of (a), or (c) a coding sequence which uses the redundancy or degeneracy of the genetic code to encode the same polypeptides. By "polypeptide" or "protein" is meant any chain of amino acids, regardless of length or post-translational modification (e.g., glycosylation or phosphorylation). Both terms are used interchangeably in the present application.

Consistent with the first aspect of the invention, amino acid sequences are provided which are homologous to any one of SEQ ID Nos: 27 to 45. As used herein, "homologous amino acid sequence" is any polypeptide which is encoded, in whole or in part, by a nucleic acid sequence which hybridizes at 25-35°C below critical melting temperature ( $T_m$ ), to any portion of the nucleic acid sequences of SEQ ID Nos: 1 to 26. A homologous amino acid sequence is one that differs from an amino acid sequence shown in any one of SEQ ID Nos: 27 to 45 by one or more amino acid substitutions. Such a sequence also encompasses serotypic variants (defined below) as well as sequences containing deletions or insertions which retain inherent characteristics of the polypeptide such as immunogenicity. Preferably, such a sequence is at least 75%, more preferably 80%, and most preferably 90% identical to any one of SEQ ID Nos: 27 to 45. Homologous amino acid sequences include sequences that are identical or substantially identical to SEQ ID Nos: 27 to 45. By "amino acid sequence substantially identical" is meant a sequence that is at least 90%, preferably 95%, more preferably 97%, and most preferably 99% identical to



an amino acid sequence of reference and that preferably differs from the sequence of reference by a majority of conservative amino acid substitutions.

Conservative amino acid substitutions are substitutions among amino acids of the same class. These classes include, for example, amino acids having uncharged polar side chains, such as asparagine, glutamine, serine, threonine, and tyrosine; amino acids having basic side chains, such as lysine, arginine, and histidine; amino acids having acidic side chains, such as aspartic acid and glutamic acid; and amino acids having nonpolar side chains, such as glycine, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan, and cysteine.

Homology is measured using sequence analysis software such as Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705. Amino acid sequences are aligned to maximize identity. Gaps may be artificially introduced into the sequence to attain proper alignment. Once the optimal alignment has been set up, the degree of homology is established by recording all of the positions in which the amino acids of both sequences are identical, relative to the total number of positions.

Homologous polynucleotide sequences are defined in a similar way. Preferably, a homologous sequence is one that is at least 45%, more preferably 60%, and most preferably 85% identical to any one of coding sequences SEQ ID Nos: 1 to 26.

Consistent with the first aspect of the invention, polypeptides having a sequence homologous to any one of SEQ ID Nos: 27 to 45 include naturally-occurring allelic variants, as well as mutants or any other non-naturally occurring variants that retain the inherent characteristics of the polypeptide of SEQ ID Nos: 27 to 45.

As is known in the art, an allelic variant is an alternate form of a polypeptide that is characterized as having a substitution, deletion, or addition of one or more amino acids that does not alter the biological function of the polypeptide.

5 By "biological function" is meant the function of the polypeptide in the cells in which it naturally occurs, even if the function is not necessary for the growth or survival of the cells. For example, the biological function of a porin is to allow the entry into cells of compounds present in the  
10 extracellular medium. Biological function is distinct from antigenic property. A polypeptide can have more than one biological function.

Allelic variants are very common in nature. For example, a bacterial species such as *C. pneumoniae*, is usually  
15 represented by a variety of strains that differ from each other by minor allelic variations. Indeed, a polypeptide that fulfills the same biological function in different strains can have an amino acid sequence (and polynucleotide sequence) that are not identical in each of the strains. Despite this  
20 variation, an immune response directed generally against many allelic variants has been demonstrated. In studies of the *Chlamydial* MOMP antigen, cross-strain antibody binding plus neutralization of infectivity occurs despite amino acid sequence variation of MOMP from strain to strain, indicating that the  
25 MOMP, when used as an immunogen, is tolerant of amino acid variations.

Polynucleotides encoding homologous polypeptides or allelic variants are retrieved by polymerase chain reaction (PCR) amplification of genomic bacterial DNA extracted by  
30 conventional methods. This involves the use of synthetic oligonucleotide primers matching upstream and downstream of the 5' and 3' ends of the encoding domain. Suitable primers are designed according to the nucleotide sequence information provided in SEQ ID Nos:1 to 26. The procedure is as follows: a

primer is selected which consists of 10 to 40, preferably 15 to 25 nucleotides. It is advantageous to select primers containing C and G nucleotides in a proportion sufficient to ensure efficient hybridization; i.e., an amount of C and G nucleotides of at least 40%, preferably 50% of the total nucleotide content.

An alternative method for retrieving polynucleotides encoding homologous polypeptides or allelic variants is by hybridization screening of a DNA or RNA library. Hybridization procedures are well-known in the art and are described in Ausubel et al., (Ref 41), Silhavy et al. (Ref 43), and Davis et al. (ref 44). Important parameters for optimizing hybridization conditions are reflected in a formula used to obtain the critical melting temperature above which two complementary DNA strands separate from each other (Ref 45). For polynucleotides of about 600 nucleotides or larger, this formula is as follows:

$$T_m = 81.5 + 0.5 \times (\% \text{ G+C}) + 1.6 \log (\text{positive ion concentration}) - 0.6 \times (\% \text{ formamide}).$$

Under appropriate stringency conditions, hybridization temperature ( $T_h$ ) is approximately 20 to 40°C, 20 to 25°C, or, preferably 30 to 40°C below the calculated  $T_m$ . Those skilled in the art will understand that optimal temperature and salt conditions can be readily determined.

For the polynucleotides of the invention, stringent conditions are achieved for both pre-hybridizing and hybridizing incubations (i) within 4-16 hours at 42°C, in 6 x SSC containing 50% formamide, or (ii) within 4-16 hours at 65°C in an aqueous 6 x SSC solution (1 M NaCl, 0.1 M sodium citrate (pH 7.0)).

Useful homologs and fragments thereof that do not occur naturally are designed using known methods for identifying regions of an antigen that are likely to tolerate amino acid sequence changes and/or deletions. As an example, homologous polypeptides from different species are compared; conserved sequences are identified. The more divergent sequences are the most likely to tolerate sequence changes. Alternatively, sequences are modified such that they become more reactive to T-

and/or B-cells. (See Table below for identification of T- and B- epitopes.) Yet another alternative is to mutate a particular amino acid residue or sequence within the polypeptide *in vitro*, then screen the mutant polypeptides for their ability to prevent 5 or treat Chlamydia infection according to the method outlined below.

A person skilled in the art will readily understand that by following the screening process of this invention, it will be determined without undue experimentation whether a particular 10 homolog of any of SEQ ID Nos: 27 to 45 may be useful in the prevention or treatment of Chlamydia infection. The screening procedure comprises the steps:

- (i) immunizing an animal, preferably mouse, with the test homolog or fragment;
- 15 (ii) inoculating the immunized animal with Chlamydia; and
- (iii) selecting those homologs or fragments which confer protection against Chlamydia.

By "conferring protection" is meant that there is a 20 reduction in severity of any of the effects of Chlamydia infection, in comparison with a control animal which was not immunized with the test homolog or fragment.

It has been previously demonstrated (Yang *et. al.*, 1993) that mice are susceptible to intranasal infection with different 25 isolates of *C. pneumoniae*. Strain AR-39 (Grayston, 1989) was used in Balb/c mice as a challenge infection model to examine the capacity of chlamydia gene products delivered as naked DNA to elicit a protective response against a sublethal *C. pneumoniae* lung infection. Protective immunity is defined as an 30 accelerated clearance of pulmonary infection.

Groups of 7 to 9 week old male Balb/c mice (6 to 10 per group) were immunized intramuscularly (i.m.) plus intranasally (i.n.) with plasmid DNA containing the coding sequence of a *C.pneumoniae* polypeptide. Saline or the plasmid vector lacking

an inserted chlamydial gene was given to groups of control animals.

For i.m. immunization alternate left and right quadriceps were injected with 100µg of DNA in 50µl of PBS on three occasions at 0, 3 and 6 weeks. For i.n. immunization, anaesthetized mice aspirated 50µl of PBS containing 50 µg DNA on three occasions at 0, 3 and 6 weeks. At week 8, immunized mice were inoculated i.n. with  $5 \times 10^5$  IFU of *C. pneumoniae*, strain AR39 in 100µl of SPG buffer to test their ability to limit the growth of a sublethal *C. pneumoniae* challenge.

Lungs were taken from mice at day 9 post-challenge and immediately homogenised in SPG buffer (7.5% sucrose, 5mM glutamate, 12.5mM phosphate pH7.5). The homogenate was stored frozen at -70°C until assay. Dilutions of the homogenate were assayed for the presence of infectious chlamydia by inoculation onto monolayers of susceptible cells. The inoculum was centrifuged onto the cells at 3000rpm for 1 hour, then the cells were incubated for three days at 35°C in the presence of 1µg/ml cycloheximide. After incubation the monolayers were fixed with formalin and methanol then immunoperoxidase stained for the presence of chlamydial inclusions using convalescent sera from rabbits infected with *C.pneumoniae* and metal-enhanced DAB as a peroxidase substrate.

Consistent with the first aspect of the invention, polypeptide derivatives are provided that are partial sequences of SEQ ID Nos: 27 to 45, partial sequences of polypeptide sequences homologous to SEQ ID Nos: 27 to 45, polypeptides derived from full-length polypeptides by internal deletion, and fusion proteins.

It is an accepted practice in the field of immunology to use fragments and variants of protein immunogens as vaccines, as all that is required to induce an immune response to a protein is a small (e.g., 8 to 10 amino acid) immunogenic region of the

protein. Various short synthetic peptides corresponding to surface-exposed antigens of pathogens other than *Chlamydia* have been shown to be effective vaccine antigens against their respective pathogens, e.g. an 11 residue peptide of murine mammary tumor virus (Ref 38), a 16-residue peptide of Semliki Forest virus (Ref 39), and two overlapping peptides of 15 residues each from canine parvovirus (Ref 40).

Accordingly, it will be readily apparent to one skilled in the art, having read the present description, that partial sequences of SEQ ID Nos: 27 to 45 or their homologous amino acid sequences are inherent to the full-length sequences and are taught by the present invention. Such polypeptide fragments preferably are at least 12 amino acids in length. Advantageously, they are at least 20 amino acids, preferably at least 50 amino acids, more preferably at least 75 amino acids, and most preferably at least 100 amino acids in length.

Polynucleotides of 30 to 600 nucleotides encoding partial sequences of sequences homologous to SEQ ID Nos: 27 to 45 are retrieved by PCR amplification using the parameters outlined above and using primers matching the sequences upstream and downstream of the 5' and 3' ends of the fragment to be amplified. The template polynucleotide for such amplification is either the full length polynucleotide homologous to one of SEQ ID Nos: 1 to 26, or a polynucleotide contained in a mixture of polynucleotides such as a DNA or RNA library. As an alternative method for retrieving the partial sequences, screening hybridization is carried out under conditions described above and using the formula for calculating  $T_m$ . Where fragments of 30 to 600 nucleotides are to be retrieved, the calculated  $T_m$  is corrected by subtracting (600/polynucleotide size in base pairs) and the stringency conditions are defined by a hybridization temperature that is 5 to 10°C below  $T_m$ . Where oligonucleotides shorter than 20-30 bases are to be obtained, the formula for calculating the  $T_m$  is as follows:  $T_m = 4 \times (G+C)$

+ 2 (A+T). For example, an 18 nucleotide fragment of 50% G+C would have an approximate  $T_m$  of 54°C. Short peptides that are fragments of SEQ. ID Nos. 27 to 45 or their homologous sequences, are obtained directly by chemical synthesis (E. Gross and H. J. Meinhofer, 4 The Peptides: Analysis, Synthesis, Biology; Modern Techniques of Peptide Synthesis, John Wiley & Sons (1981), and M. Bodanzki, Principles of Peptide Synthesis, Springer -Verlag (1984)).

Useful polypeptide derivatives, e.g., polypeptide fragments, are designed using computer-assisted analysis of amino acid sequences. This identifies probable surface-exposed, antigenic regions (Ref 37). An analysis of the 13 amino acid sequences contained in SEQ ID Nos: 27 to 45, based on the product of flexibility and hydrophobicity propensities using the program SEQSEE (Wishart DS, et al. "SEQSEE: a comprehensive program suite for protein sequence analysis." *Comput Appl Biosci.* 1994 Apr;10(2):121-32), reveal a number of potential B- and T-cell epitopes which may be used as a basis for selecting useful immunogenic fragments and variants. The results are shown in Figures 27 to 39. This analysis uses a reasonable combination of external surface features that is likely to be recognized by antibodies. Probable T-cell epitopes for HLA-A0201 MHC subclass were revealed by an algorithm written at Connaught Laboratories that emulates an approach developed at the NIH (Parker KC, et al. "Peptide binding to MHC class I molecules: implications for antigenic peptide prediction." *Immunol Res* 1995;14(1):34-57 ).

Epitopes which induce a protective T cell-dependent immune response are present throughout the length of the polypeptide. However, some epitopes may be masked by secondary and tertiary structures of the polypeptide. To reveal such masked epitopes large internal deletions are created which remove much of the original protein structure and exposes the masked epitopes. Such internal deletions sometimes effects the

additional advantage of removing immunodominant regions of high variability among strains. Polynucleotides encoding polypeptide fragments and polypeptides having large internal deletions are constructed using standard methods (Ref 41). Such methods  
5 include standard PCR, inverse PCR, restriction enzyme treatment of cloned DNA molecules, or the method of Kunkel et al. (Ref 42). Components for these methods and instructions for their use are readily available from various commercial sources such as Stratagene. Once the deletion mutants have been constructed,  
10 they are tested for their ability to prevent or treat Chlamydia infection as described above.

As used herein, a fusion polypeptide is one that contains a polypeptide or a polypeptide derivative of the invention fused at the N- or C-terminal end to any other polypeptide  
15 (hereinafter referred to as a peptide tail). A simple way to obtain such a fusion polypeptide is by translation of an in-frame fusion of the polynucleotide sequences, i.e., a hybrid gene. The hybrid gene encoding the fusion polypeptide is inserted into an expression vector which is used to transform or  
20 transfect a host cell. Alternatively, the polynucleotide sequence encoding the polypeptide or polypeptide derivative is inserted into an expression vector in which the polynucleotide encoding the peptide tail is already present. Such vectors and instructions for their use are commercially available, e.g.  
25 the pMal-c2 or pMal-p2 system from New England Biolabs, in which the peptide tail is a maltose binding protein, the glutathione-S-transferase system of Pharmacia, or the His-Tag system available from Novagen. These and other expression systems provide convenient means for further purification of  
30 polypeptides and derivatives of the invention.

An advantageous example of a fusion polypeptide is one where the polypeptide or homolog or fragment of the invention is fused to a polypeptide having adjuvant activity, such as subunit B of either cholera toxin or *E. coli* heat-labile toxin. Another



advantageous fusion is one where the polypeptide, homolog or fragment is fused to a strong T-cell epitope or B-cell epitope. Such an epitope may be one known in the art (e.g. the Hepatitis B virus core antigen, D.R. Millich et al., "Antibody production to the nucleocapsid and envelope of the Hepatitis B virus primed by a single synthetic T cell site", Nature. 1987. 329:547-549), or one which has been identified in another polypeptide of the invention (Table ). Consistent with this aspect of the invention is a fusion polypeptide comprising T- or B-cell epitopes from one of SEQ ID Nos: 27 to 45 or its homolog or fragment, wherein the epitopes are derived from multiple variants of said polypeptide or homolog or fragment, each variant differing from another in the location and sequence of its epitope within the polypeptide. Such a fusion is effective in the prevention and treatment of Chlamydia infection since it optimizes the T- and B-cell response to the overall polypeptide, homolog or fragment.

To effect fusion, the polypeptide of the invention is fused to the N-, or preferably, to the C-terminal end of the polypeptide having adjuvant activity or T- or B-cell epitope. Alternatively, a polypeptide fragment of the invention is inserted internally within the amino acid sequence of the polypeptide having adjuvant activity. The T- or B-cell epitope may also be inserted internally within the amino acid sequence of the polypeptide of the invention.

Consistent with the first aspect, the polynucleotides of the invention also encode hybrid precursor polypeptides containing heterologous signal peptides, which mature into polypeptides of the invention. By "heterologous signal peptide" is meant a signal peptide that is not found in naturally-occurring precursors of polypeptides of the invention.

A polynucleotide molecule according to the invention, including RNA, DNA, or modifications or combinations thereof, have various applications. A DNA molecule is used, for example,

(i) in a process for producing the encoded polypeptide in a recombinant host system, (ii) in the construction of vaccine vectors such as poxviruses, which are further used in methods and compositions for preventing and/or treating *Chlamydia* infection, (iii) as a vaccine agent (as well as an RNA molecule), in a naked form or formulated with a delivery vehicle and, (iv) in the construction of attenuated *Chlamydia* strains that can over-express a polynucleotide of the invention or express it in a non-toxic, mutated form.

10           Accordingly, a second aspect of the invention encompasses (i) an expression cassette containing a DNA molecule of the invention placed under the control of the elements required for expression, in particular under the control of an appropriate promoter; (ii) an expression vector containing an expression  
15 cassette of the invention; (iii) a procaryotic or eucaryotic cell transformed or transfected with an expression cassette and/or vector of the invention, as well as (iv) a process for producing a polypeptide or polypeptide derivative encoded by a polynucleotide of the invention, which involves culturing a  
20 procaryotic or eucaryotic cell transformed or transfected with an expression cassette and/or vector of the invention, under conditions that allow expression of the DNA molecule of the invention and, recovering the encoded polypeptide or polypeptide derivative from the cell culture.

25           A recombinant expression system is selected from procaryotic and eucaryotic hosts. Eucaryotic hosts include yeast cells (e.g., *Saccharomyces cerevisiae* or *Pichia pastoris*), mammalian cells (e.g., COS1, NIH3T3, or JEG3 cells), arthropods cells (e.g., *Spodoptera frugiperda* (SF9) cells), and plant  
30 cells. A preferred expression system is a procaryotic host such as *E. coli*. Bacterial and eucaryotic cells are available from a number of different sources including commercial sources to those skilled in the art, e.g., the American Type Culture Collection (ATCC; Rockville, Maryland). Commercial sources of

cells used for recombinant protein expression also provide instructions for usage of the cells.

The choice of the expression system depends on the features desired for the expressed polypeptide. For example, it  
5 may be useful to produce a polypeptide of the invention in a particular lipidated form or any other form.

One skilled in the art would readily understand that not all vectors and expression control sequences and hosts would be expected to express equally well the polynucleotides of this  
10 invention. With the guidelines described below, however, a selection of vectors, expression control sequences and hosts may be made without undue experimentation and without departing from the scope of this invention.

In selecting a vector, the host must be chosen that is  
15 compatible with the vector which is to exist and possibly replicate in it. Considerations are made with respect to the vector copy number, the ability to control the copy number, expression of other proteins such as antibiotic resistance. In selecting an expression control sequence, a number of variables  
20 are considered. Among the important variable are the relative strength of the sequence (e.g. the ability to drive expression under various conditions), the ability to control the sequence's function, compatibility between the polynucleotide to be expressed and the control sequence (e.g. secondary structures  
25 are considered to avoid hairpin structures which prevent efficient transcription). In selecting the host, unicellular hosts are selected which are compatible with the selected vector, tolerant of any possible toxic effects of the expressed product, able to secrete the expressed product efficiently if  
30 such is desired, to be able to express the product in the desired conformation, to be easily scaled up, and to which ease of purification of the final product.

The choice of the expression cassette depends on the host system selected as well as the features desired for the

expressed polypeptide. Typically, an expression cassette includes a promoter that is functional in the selected host system and can be constitutive or inducible; a ribosome binding site; a start codon (ATG) if necessary; a region encoding a  
5 signal peptide, e.g., a lipidation signal peptide; a DNA molecule of the invention; a stop codon; and optionally a 3' terminal region (translation and/or transcription terminator). The signal peptide encoding region is adjacent to the polynucleotide of the invention and placed in proper reading  
10 frame. The signal peptide-encoding region is homologous or heterologous to the DNA molecule encoding the mature polypeptide and is compatible with the secretion apparatus of the host used for expression. The open reading frame constituted by the DNA molecule of the invention, solely or together with the signal  
15 peptide, is placed under the control of the promoter so that transcription and translation occur in the host system. Promoters and signal peptide encoding regions are widely known and available to those skilled in the art and include, for example, the promoter of *Salmonella typhimurium* (and  
20 derivatives) that is inducible by arabinose (promoter araB) and is functional in Gram-negative bacteria such as *E. coli* (as described in U.S. Patent No. 5,028,530 and in Cagnon et al., (Ref 46)); the promoter of the gene of bacteriophage T7 encoding RNA polymerase, that is functional in a number of *E. coli*  
25 strains expressing T7 polymerase (described in U.S. Patent No. 4,952,496); OspA lipidation signal peptide ; and RlpB lipidation signal peptide (Ref 47).

The expression cassette is typically part of an expression vector, which is selected for its ability to  
30 replicate in the chosen expression system. Expression vectors (e.g., plasmids or viral vectors) can be chosen, for example, from those described in Pouwels et al. (Cloning Vectors: A Laboratory Manual 1985, Supp. 1987). Suitable expression vectors can be purchased from various commercial sources.

Methods for transforming/transfecting host cells with expression vectors are well-known in the art and depend on the host system selected as described in Ausubel et al., (Ref 41).

Upon expression, a recombinant polypeptide of the invention (or a polypeptide derivative) is produced and remains in the intracellular compartment, is secreted/excreted in the extracellular medium or in the periplasmic space, or is embedded in the cellular membrane. The polypeptide is recovered in a substantially purified form from the cell extract or from the supernatant after centrifugation of the recombinant cell culture. Typically, the recombinant polypeptide is purified by antibody-based affinity purification or by other well-known methods that can be readily adapted by a person skilled in the art, such as fusion of the polynucleotide encoding the polypeptide or its derivative to a small affinity binding domain. Antibodies useful for purifying by immunoaffinity the polypeptides of the invention are obtained as described below.

A polynucleotide of the invention can also be useful as a vaccine. There are two major routes, either using a viral or bacterial host as gene delivery vehicle (live vaccine vector) or administering the gene in a free form, e.g., inserted into a plasmid. Therapeutic or prophylactic efficacy of a polynucleotide of the invention is evaluated as described below.

Accordingly, a third aspect of the invention provides (i) a vaccine vector such as a poxvirus, containing a DNA molecule of the invention, placed under the control of elements required for expression; (ii) a composition of matter comprising a vaccine vector of the invention, together with a diluent or carrier; specifically (iii) a pharmaceutical composition containing a therapeutically or prophylactically effective amount of a vaccine vector of the invention; (iv) a method for inducing an immune response against *Chlamydia* in a mammal (e.g., a human; alternatively, the method can be used in veterinary applications for treating or preventing *Chlamydia* infection of

animals, e.g., cats or birds), which involves administering to the mammal an immunogenically effective amount of a vaccine vector of the invention to elicit a protective or therapeutic immune response to *Chlamydia* ; and particularly, (v) a method  
5 for preventing and/or treating a *Chlamydia* (e.g., *C. trachomatis*, *C. psittaci*, *C. pneumonia*, *C. pecorum*) infection, which involves administering a prophylactic or therapeutic amount of a vaccine vector of the invention to an infected individual. Additionally, the third aspect of the invention  
10 encompasses the use of a vaccine vector of the invention in the preparation of a medicament for preventing and/or treating *Chlamydia* infection.

As used herein, a vaccine vector expresses one or several polypeptides or derivatives of the invention, as well as at  
15 least one additional *Chlamydia* antigen (??), fragment, homolog, mutant, or derivative thereof. The vaccine vector may express additionally a cytokine, such as interleukin-2 (IL-2) or interleukin-12 (IL-12), that enhances the immune response (adjuvant effect). It is understood that each of the components  
20 to be expressed is placed under the control of elements required for expression in a mammalian cell.

Consistent with the third aspect of the invention is a composition comprising several vaccine vectors, each of them capable of expressing a polypeptide or derivative of the  
25 invention. A composition may also comprise a vaccine vector capable of expressing an additional *Chlamydia* antigen, or a subunit, fragment, homolog, mutant, or derivative thereof; or a cytokine such as IL-2 or IL-12.

Vaccination methods for treating or preventing infection  
30 in a mammal comprises use of a vaccine vector of the invention to be administered by any conventional route , particularly to a mucosal (e.g., ocular, intranasal, oral, gastric, pulmonary, intestinal, rectal, vaginal, or urinary tract) surface or via the parenteral (e.g., subcutaneous, intradermal, intramuscular,

intravenous, or intraperitoneal) route. Preferred routes depend upon the choice of the vaccine vector. Treatment may be effected in a single dose or repeated at intervals. The appropriate dosage depends on various parameters understood by 5 skilled artisans such as the vaccine vector itself, the route of administration or the condition of the mammal to be vaccinated (weight, age and the like).

Live vaccine vectors available in the art include viral vectors such as adenoviruses and poxviruses as well as bacterial 10 vectors, e.g., *Shigella*, *Salmonella*, *Vibrio cholerae*, *Lactobacillus*, Bacille bilié de Calmette-Guérin (BCG), and *Streptococcus*.

An example of an adenovirus vector, as well as a method for constructing an adenovirus vector capable of expressing a 15 DNA molecule of the invention, are described in U.S. Patent No. 4,920,209. Poxvirus vectors include vaccinia and canary pox virus, described in U.S. Patent No. 4,722,848 and U.S. Patent No. 5,364,773, respectively. (Also see, e.g., Tartaglia et al., Virology (1992) 188:217) for a description of a vaccinia virus 20 vector and Taylor et al, Vaccine (1995) 13:539 for a reference of a canary pox.) Poxvirus vectors capable of expressing a polynucleotide of the invention are obtained by homologous recombination as described in Kieny et al., Nature (1984) 312:163 so that the polynucleotide of the invention is inserted 25 in the viral genome under appropriate conditions for expression in mammalian cells. Generally, the dose of vaccine viral vector, for therapeutic or prophylactic use, can be of from about  $1 \times 10^4$  to about  $1 \times 10^{11}$ , advantageously from about  $1 \times 10^7$  to about  $1 \times 10^{10}$ , preferably of from about  $1 \times 10^7$  to about  $1 \times 10^9$  30 plaque-forming units per kilogram. Preferably, viral vectors are administered parenterally; for example, in 3 doses, 4 weeks apart. It is preferable to avoid adding a chemical adjuvant to a composition containing a viral vector of the invention and

thereby minimizing the immune response to the viral vector itself.

Non-toxicogenic *Vibrio cholerae* mutant strains that are useful as a live oral vaccine are known. Mekalanos et al., Nature (1983) 306:551 and U.S. Patent No. 4,882,278 describe strains which have a substantial amount of the coding sequence of each of the two *ctxA* alleles deleted so that no functional *cholerae* toxin is produced. WO 92/11354 describes a strain in which the *irgA* locus is inactivated by mutation; this mutation can be combined in a single strain with *ctxA* mutations. WO 94/1533 describes a deletion mutant lacking functional *ctxA* and *attRS1* DNA sequences. These mutant strains are genetically engineered to express heterologous antigens, as described in WO 94/19482. An effective vaccine dose of a *Vibrio cholerae* strain capable of expressing a polypeptide or polypeptide derivative encoded by a DNA molecule of the invention contains about  $1 \times 10^5$  to about  $1 \times 10^9$ , preferably about  $1 \times 10^6$  to about  $1 \times 10^8$ , viable bacteria in a volume appropriate for the selected route of administration. Preferred routes of administration include all mucosal routes; most preferably, these vectors are administered intranasally or orally.

Attenuated *Salmonella typhimurium* strains, genetically engineered for recombinant expression of heterologous antigens or not, and their use as oral vaccines are described in Nakayama et al. (Bio/Technology (1988) 6:693) and WO 92/11361. Preferred routes of administration include all mucosal routes; most preferably, these vectors are administered intranasally or orally.

Other bacterial strains used as vaccine vectors in the context of the present invention are described in High et al., EMBO (1992) 11:1991 and Sizemore et al., Science (1995) 270:299 (*Shigella flexneri*); Medaglini et al., Proc. Natl. Acad. Sci. USA (1995) 92:6868 (*Streptococcus gordonii*), Flynn J.L., Cell.



Mol. Biol. (1994) 40 (suppl. I):31, WO 88/6626, WO 90/0594, WO 91/13157, WO 92/1796, and WO 92/21376 (Bacille Calmette Guerin).

In bacterial vectors, the polynucleotide of the invention is inserted into the bacterial genome or remains in a free state as part of a plasmid.

The composition comprising a vaccine bacterial vector of the present invention may further contain an adjuvant. A number of adjuvants are known to those skilled in the art. Preferred adjuvants are selected from the list provided below.

10 Accordingly, a fourth aspect of the invention provides (i) a composition of matter comprising a polynucleotide of the invention, together with a diluent or carrier; (ii) a pharmaceutical composition comprising a therapeutically or prophylactically effective amount of a polynucleotide of the invention; (iii) a method for inducing an immune response against *Chlamydia* in a mammal by administration of an immunogenically effective amount of a polynucleotide of the invention to elicit a protective immune response to *Chlamydia*; and particularly, (iv) a method for preventing and/or treating a  
20 *Chlamydia* (e.g., *C. trachomatis*, *C. psittaci*, *C. pneumoniae*, or *C. pecorum*) infection, by administering a prophylactic or therapeutic amount of a polynucleotide of the invention to an infected individual. Additionally, the fourth aspect of the invention encompasses the use of a polynucleotide of the  
25 invention in the preparation of a medicament for preventing and/or treating *Chlamydia* infection. A preferred use includes the use of a DNA molecule placed under conditions for expression in a mammalian cell, especially in a plasmid that is unable to replicate in mammalian cells and to substantially integrate in a  
30 mammalian genome.

Use of the polynucleotides of the invention include their administration to a mammal as a vaccine, for therapeutic or prophylactic purposes. Such polynucleotides are used in the form of DNA as part of a plasmid that is unable to replicate in

a mammalian cell and unable to integrate into the mammalian genome. Typically, such a DNA molecule is placed under the control of a promoter suitable for expression in a mammalian cell. The promoter functions either ubiquitously or tissue-specifically. Examples of non-tissue specific promoters include the early Cytomegalovirus (CMV) promoter (described in U.S. Patent No. 4,168,062) and the Rous Sarcoma Virus promoter (described in Norton & Coffin, *Molec. Cell Biol.* (1985) 5:281). An example of a tissue-specific promoter is the desmin promoter which drives expression in muscle cells (Li et al., *Gene* (1989) 78:243, Li & Paulin, *J. Biol. Chem.* (1991) 266:6562 and Li & Paulin, *J. Biol. Chem.* (1993) 268:10403). Use of promoters is well-known to those skilled in the art. Useful vectors are described in numerous publications, specifically WO 94/21797 and Hartikka et al., *Human Gene Therapy* (1996) 7:1205.

Polynucleotides of the invention which are used as a vaccine encode either a precursor or a mature form of the corresponding polypeptide. In the precursor form, the signal peptide is either homologous or heterologous. In the latter case, a eucaryotic leader sequence such as the leader sequence of the tissue-type plasminogen factor (tPA) is preferred.

As used herein, a composition of the invention contains one or several polynucleotides with optionally at least one additional polynucleotide encoding another *Chlamydia* antigen such as urease subunit A, B, or both, or a fragment, derivative, mutant, or analog thereof. The composition may also contain an additional polynucleotide encoding a cytokine, such as interleukin-2 (IL-2) or interleukin-12 (IL-12) so that the immune response is enhanced. These additional polynucleotides are placed under appropriate control for expression. Advantageously, DNA molecules of the invention and/or additional DNA molecules to be included in the same composition, are present in the same plasmid.

Standard techniques of molecular biology for preparing and purifying polynucleotides are used in the preparation of polynucleotide therapeutics of the invention. For use as a vaccine, a polynucleotide of the invention is formulated according to various methods outlined below.

One method utilizes the polynucleotide in a naked form, free of any delivery vehicles. Such a polynucleotide is simply diluted in a physiologically acceptable solution such as sterile saline or sterile buffered saline, with or without a carrier. When present, the carrier preferably is isotonic, hypotonic, or weakly hypertonic, and has a relatively low ionic strength, such as provided by a sucrose solution, e.g., a solution containing 20% sucrose.

An alternative method utilizes the polynucleotide in association with agents that assist in cellular uptake. Examples of such agents are (i) chemicals that modify cellular permeability, such as bupivacaine (see, e.g., WO 94/16737), (ii) liposomes for encapsulation of the polynucleotide, or (iii) cationic lipids or silica, gold, or tungsten microparticles which associate themselves with the polynucleotides.

Anionic and neutral liposomes are well-known in the art (see, e.g., Liposomes: A Practical Approach, RPC New Ed, IRL press (1990), for a detailed description of methods for making liposomes) and are useful for delivering a large range of products, including polynucleotides. Cationic lipids are also known in the art and are commonly used for gene delivery. Such lipids include Lipofectin<sup>TM</sup> also known as DOTMA (N-[1-(2,3-dioleyloxy)propyl]-N,N,N-trimethylammonium chloride), DOTAP (1,2-bis(oleyloxy)-3-(trimethylammonio)propane), DDAB (dimethyldioctadecylammonium bromide), DOGS (dioctadecylamidoglycyl spermine) and cholesterol derivatives such as DC-Chol (3 beta-(N-(N',N'-dimethyl aminomethane)-carbamoyl) cholesterol). A description of these cationic lipids

can be found in EP 187,702, WO 90/11092, U.S. Patent No. 5,283,185, WO 91/15501, WO 95/26356, and U.S. Patent No. 5,527,928. Cationic lipids for gene delivery are preferably used in association with a neutral lipid such as DOPE (dioleyl 5 phosphatidylethanolamine), as described in WO 90/11092 as an example.

Formulation containing cationic liposomes may optionally contain other transfection-facilitating compounds. A number of them are described in WO 93/18759, WO 93/19768, WO 94/25608, and 10 WO 95/2397. They include spermine derivatives useful for facilitating the transport of DNA through the nuclear membrane (see, for example, WO 93/18759) and membrane-permeabilizing compounds such as GALA, Gramicidine S, and cationic bile salts (see, for example, WO 93/19768).

15 Gold or tungsten microparticles are used for gene delivery, as described in WO 91/359, WO 93/17706, and Tang et al. (Nature (1992) 356:152). The microparticle-coated polynucleotide is injected via intradermal or intraepidermal routes using a needleless injection device ("gene gun"), such as 20 those described in U.S. Patent No. 4,945,050, U.S. Patent No. 5,015,580, and WO 94/24263.

The amount of DNA to be used in a vaccine recipient depends, e.g., on the strength of the promoter used in the DNA construct, the immunogenicity of the expressed gene product, the 25 condition of the mammal intended for administration (e.g., the weight, age, and general health of the mammal), the mode of administration, and the type of formulation. In general, a therapeutically or prophylactically effective dose from about 1 µg to about 1 mg, preferably, from about 10 µg to about 800 µg 30 and, more preferably, from about 25 µg to about 250 µg, can be administered to human adults. The administration can be achieved in a single dose or repeated at intervals.

The route of administration is any conventional route used in the vaccine field. As general guidance, a

polynucleotide of the invention is administered via a mucosal surface, e.g., an ocular, intranasal, pulmonary, oral, intestinal, rectal, vaginal, and urinary tract surface; or via a parenteral route, e.g., by an intravenous, subcutaneous, 5 intraperitoneal, intradermal, intraepidermal, or intramuscular route. The choice of administration route depends on the formulation that is selected. A polynucleotide formulated in association with bupivacaine is advantageously administered into muscles. When a neutral or anionic liposome or a cationic 10 lipid, such as DOTMA or DC-Chol, is used, the formulation can be advantageously injected via intravenous, intranasal (aerosolization), intramuscular, intradermal, and subcutaneous routes. A polynucleotide in a naked form can advantageously be administered via the intramuscular, intradermal, or sub- 15 cutaneous routes.

Although not absolutely required, such a composition can also contain an adjuvant. If so, a systemic adjuvant that does not require concomitant administration in order to exhibit an adjuvant effect is preferable such as, e.g., QS21, which is 20 described in U.S. Patent No. 5,057,546.

The sequence information provided in the present application enables the design of specific nucleotide probes and primers that are used for diagnostic purposes. Accordingly, a fifth aspect of the invention provides a nucleotide probe or 25 primer having a sequence found in or derived by degeneracy of the genetic code from a sequence shown in any one of SEQ ID Nos:1 to 26.

The term "probe" as used in the present application refers to DNA (preferably single stranded) or RNA molecules (or 30 modifications or combinations thereof) that hybridize under the stringent conditions, as defined above, to nucleic acid molecules having SEQ ID Nos: 1 to 26 or to sequences homologous to SEQ ID Nos:1 to 26, or to their complementary or anti-sense sequences. Generally, probes are significantly shorter than

full-length sequences . Such probes contain from about 5 to about 100, preferably from about 10 to about 80, nucleotides. In particular, probes have sequences that are at least 75%, preferably at least 85%, more preferably 95% homologous to a portion of any of SEQ ID Nos:1 to 26 or that are complementary to such sequences. Probes may contain modified bases such as inosine, methyl-5-deoxycytidine, deoxyuridine, dimethylamino-5-deoxyuridine, or diamino-2, 6-purine. Sugar or phosphate residues may also be modified or substituted. For example, a deoxyribose residue may be replaced by a polyamide (Nielsen et al., Science (1991) 254:1497) and phosphate residues may be replaced by ester groups such as diphosphate, alkyl, arylphosphonate and phosphorothioate esters. In addition, the 2'-hydroxyl group on ribonucleotides may be modified by including such groups as alkyl groups.

Probes of the invention are used in diagnostic tests, as capture or detection probes. Such capture probes are conventionally immobilized on a solid support, directly or indirectly, by covalent means or by passive adsorption. A detection probe is labelled by a detection marker selected from: radioactive isotopes, enzymes such as peroxidase, alkaline phosphatase, and enzymes able to hydrolyze a chromogenic, fluorogenic, or luminescent substrate, compounds that are chromogenic, fluorogenic, or luminescent, nucleotide base analogs, and biotin.

Probes of the invention are used in any conventional hybridization technique, such as dot blot (Maniatis et al., Molecular Cloning: A Laboratory Manual (1982) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York), Southern blot (Southern, J. Mol. Biol. (1975) 98:503), northern blot (identical to Southern blot with the exception that RNA is used as a target), or the sandwich technique (Dunn et al., Cell (1977) 12:23). The latter technique involves the use of a specific capture probe and/or a specific detection probe with

nucleotide sequences that at least partially differ from each other.

A primer is a probe of usually about 10 to about 40 nucleotides that is used to initiate enzymatic polymerization of DNA in an amplification process (e.g., PCR), in an elongation process, or in a reverse transcription method. Primers used in diagnostic methods involving PCR are labeled by methods known in the art.

As described herein, the invention also encompasses (i) a reagent comprising a probe of the invention for detecting and/or identifying the presence of *Chlamydia* in a biological material; (ii) a method for detecting and/or identifying the presence of *Chlamydia* in a biological material, in which (a) a sample is recovered or derived from the biological material, (b) DNA or RNA is extracted from the material and denatured, and (c) exposed to a probe of the invention, for example, a capture, detection probe or both, under stringent hybridization conditions, such that hybridization is detected; and (iii) a method for detecting and/or identifying the presence of *Chlamydia* in a biological material, in which (a) a sample is recovered or derived from the biological material, (b) DNA is extracted therefrom, (c) the extracted DNA is primed with at least one, and preferably two, primers of the invention and amplified by polymerase chain reaction, and (d) the amplified DNA fragment is produced.

It is apparent that disclosure of polynucleotide sequences of SEQ ID Nos: 1 to 26, their homolog, and partial sequences of either enable their corresponding amino acid sequences. Accordingly, a sixth aspect of the invention features a substantially purified polypeptide or polypeptide derivative having an amino acid sequence encoded by a polynucleotide of the invention.

A "substantially purified polypeptide" as used herein is defined as a polypeptide that is separated from the environment

in which it naturally occurs and/or that is free of the majority of the polypeptides that are present in the environment in which it was synthesized. For example, a substantially purified polypeptide is free from cytoplasmic polypeptides. Those  
5 skilled in the art would readily understand that the polypeptides of the invention may be purified from a natural source, *i.e.*, a *Chlamydia* strain, or produced by recombinant means.

Consistent with the sixth aspect of the invention are  
10 polypeptides, homologs or fragments which are modified or treated to enhance their immunogenicity in the target animal, in whom the polypeptide, homolog or fragments are intended to confer protection against *Chlamydia*. Such modifications or treatments include: amino acid substitutions with an amino acid  
15 derivative such as 3-methylhistidine, 4-hydroxyproline, 5-hydroxylysine etc., modifications or deletions which are carried out after preparation of the polypeptide, homolog or fragment, such as the modification of free amino, carboxyl or hydroxyl side groups of the amino acids.

20 Identification of homologous polypeptides or polypeptide derivatives encoded by polynucleotides of the invention which have specific antigenicity is achieved by screening for cross-reactivity with an antiserum raised against the polypeptide of reference having an amino acid sequence of any one of SEQ ID  
25 Nos: 27 to 45. The procedure is as follows: a monospecific hyperimmune antiserum is raised against a purified reference polypeptide, a fusion polypeptide (for example, an expression product of MBP, GST, or His-tag systems), or a synthetic peptide predicted to be antigenic. Where an antiserum is raised  
30 against a fusion polypeptide, two different fusion systems are employed. Specific antigenicity can be determined according to a number of methods, including Western blot (Towbin et al., Proc. Natl. Acad. Sci. USA (1979) 76:4350), dot blot, and ELISA, as described below.



In a Western blot assay, the product to be screened, either as a purified preparation or a total *E. coli* extract, is submitted to SDS-Page electrophoresis as described by Laemmli (Nature (1970) 227:680). After transfer to a nitrocellulose  
5 membrane, the material is further incubated with the monospecific hyperimmune antiserum diluted in the range of dilutions from about 1:5 to about 1:5000, preferably from about 1:100 to about 1:500. Specific antigenicity is shown once a band corresponding to the product exhibits reactivity at any of  
10 the dilutions in the above range.

In an ELISA assay, the product to be screened is preferably used as the coating antigen. A purified preparation is preferred, although a whole cell extract can also be used. Briefly, about 100  $\mu$ l of a preparation at about 10  $\mu$ g protein/ml  
15 are distributed into wells of a 96-well polycarbonate ELISA plate. The plate is incubated for 2 hours at 37°C then overnight at 4°C. The plate is washed with phosphate buffer saline (PBS) containing 0.05% Tween 20 (PBS/Tween buffer). The wells are saturated with 250  $\mu$ l PBS containing 1% bovine serum  
20 albumin (BSA) to prevent non-specific antibody binding. After 1 hour incubation at 37°C, the plate is washed with PBS/Tween buffer. The antiserum is serially diluted in PBS/Tween buffer containing 0.5% BSA. 100  $\mu$ l of dilutions are added per well. The plate is incubated for 90 minutes at 37°C, washed and  
25 evaluated according to standard procedures. For example, a goat anti-rabbit peroxidase conjugate is added to the wells when specific antibodies were raised in rabbits. Incubation is carried out for 90 minutes at 37°C and the plate is washed. The reaction is developed with the appropriate substrate and the  
30 reaction is measured by colorimetry (absorbance measured spectrophotometrically). Under the above experimental conditions, a positive reaction is shown by O.D. values greater than a non immune control serum.

In a dot blot assay, a purified product is preferred, although a whole cell extract can also be used. Briefly, a solution of the product at about 100 µg/ml is serially two-fold diluted in 50 mM Tris-HCl (pH 7.5). 100 µl of each dilution are  
5 applied to a nitrocellulose membrane 0.45 µm set in a 96-well dot blot apparatus (Biorad). The buffer is removed by applying vacuum to the system. Wells are washed by addition of 50 mM Tris-HCl (pH 7.5) and the membrane is air-dried. The membrane is saturated in blocking buffer (50 mM Tris-HCl (pH 7.5) 0.15 M  
10 NaCl, 10 g/L skim milk) and incubated with an antiserum dilution from about 1:50 to about 1:5000, preferably about 1:500. The reaction is revealed according to standard procedures. For example, a goat anti-rabbit peroxidase conjugate is added to the wells when rabbit antibodies are used. Incubation is carried  
15 out 90 minutes at 37°C and the blot is washed. The reaction is developed with the appropriate substrate and stopped. The reaction is measured visually by the appearance of a colored spot, e.g., by colorimetry. Under the above experimental conditions, a positive reaction is shown once a colored spot is  
20 associated with a dilution of at least about 1:5, preferably of at least about 1:500.

Therapeutic or prophylactic efficacy of a polypeptide or derivative of the invention can be evaluated as described below. A seventh aspect of the invention provides (i) a composition of  
25 matter comprising a polypeptide of the invention together with a diluent or carrier; specifically (ii) a pharmaceutical composition containing a therapeutically or prophylactically effective amount of a polypeptide of the invention; (iii) a method for inducing an immune response against *Chlamydia* in a  
30 mammal, by administering to the mammal an immunogenically effective amount of a polypeptide of the invention to elicit a protective immune response to *Chlamydia*; and particularly, (iv) a method for preventing and/or treating a *Chlamydia* (e.g., *C. trachomatis*, *C. psittaci*, *C. pneumoniae*, or *C. pecorum*)

infection, by administering a prophylactic or therapeutic amount of a polypeptide of the invention to an infected individual. Additionally, the seventh aspect of the invention encompasses the use of a polypeptide of the invention in the preparation of  
5 a medicament for preventing and/or treating *Chlamydia* infection.

As used herein, the immunogenic compositions of the invention are administered by conventional routes known the vaccine field, in particular to a mucosal (e.g., ocular, intranasal, pulmonary, oral, gastric, intestinal, rectal,  
10 vaginal, or urinary tract) surface or via the parenteral (e.g., subcutaneous, intradermal, intramuscular, intravenous, or intraperitoneal) route. The choice of administration route depends upon a number of parameters, such as the adjuvant associated with the polypeptide. If a mucosal adjuvant is used,  
15 the intranasal or oral route is preferred. If a lipid formulation or an aluminum compound is used, the parenteral route is preferred with the sub-cutaneous or intramuscular route being most preferred. The choice also depends upon the nature of the vaccine agent. For example, a polypeptide of the  
20 invention fused to CTB or LTB is best administered to a mucosal surface.

As used herein, the composition of the invention contains one or several polypeptides or derivatives of the invention. The composition optionally contains at least one additional  
25 *Chlamydia* antigen, or a subunit, fragment, homolog, mutant, or derivative thereof.

For use in a composition of the invention, a polypeptide or derivative thereof is formulated into or with liposomes, preferably neutral or anionic liposomes, microspheres, ISCOMS,  
30 or virus-like-particles (VLPs) to facilitate delivery and/or enhance the immune response. These compounds are readily available to one skilled in the art; for example, see Liposomes: A Practical Approach (*supra*).

Adjuvants other than liposomes and the like are also used and are known in the art. Adjuvants may protect the antigen from rapid dispersal by sequestering it in a local deposit, or they may contain substances that stimulate the host to secrete  
5 factors that are chemotactic for macrophages and other components of the immune system. An appropriate selection can conventionally be made by those skilled in the art, for example, from those described below.

Treatment is achieved in a single dose or repeated as  
10 necessary at intervals, as can be determined readily by one skilled in the art. For example, a priming dose is followed by three booster doses at weekly or monthly intervals. An appropriate dose depends on various parameters including the recipient (e.g., adult or infant), the particular vaccine  
15 antigen, the route and frequency of administration, the presence/absence or type of adjuvant, and the desired effect (e.g., protection and/or treatment), as can be determined by one skilled in the art. In general, a vaccine antigen of the invention is administered by a mucosal route in an amount from  
20 about 10  $\mu$ g to about 500 mg, preferably from about 1 mg to about 200 mg. For the parenteral route of administration, the dose usually does not exceed about 1 mg, preferably about 100  $\mu$ g.

When used as vaccine agents, polynucleotides and polypeptides of the invention may be used sequentially as part  
25 of a multistep immunization process. For example, a mammal is initially primed with a vaccine vector of the invention such as a pox virus, e.g., via the parenteral route, and then boosted twice with the polypeptide encoded by the vaccine vector, e.g., via the mucosal route. In another example, liposomes associated  
30 with a polypeptide or derivative of the invention is also used for priming, with boosting being carried out mucosally using a soluble polypeptide or derivative of the invention in combination with a mucosal adjuvant (e.g., LT).

A polypeptide derivative of the invention is also used in accordance with the seventh aspect as a diagnostic reagent for detecting the presence of anti-*Chlamydia* antibodies, e.g., in a blood sample. Such polypeptides are about 5 to about 80,  
5 preferably about 10 to about 50 amino acids in length. They are either labeled or unlabeled, depending upon the diagnostic method. Diagnostic methods involving such a reagent are described below.

Upon expression of a DNA molecule of the invention, a  
10 polypeptide or polypeptide derivative is produced and purified using known laboratory techniques. As described above, the polypeptide or polypeptide derivative may be produced as a fusion protein containing a fused tail that facilitates purification. The fusion product is used to immunize a small  
15 mammal, e.g., a mouse or a rabbit, in order to raise antibodies against the polypeptide or polypeptide derivative (monospecific antibodies). Accordingly, an eighth aspect of the invention provides a monospecific antibody that binds to a polypeptide or polypeptide derivative of the invention.

20 By "monospecific antibody" is meant an antibody that is capable of reacting with a unique naturally-occurring *Chlamydia* polypeptide. An antibody of the invention is either polyclonal or monoclonal. Monospecific antibodies may be recombinant, e.g., chimeric (e.g., constituted by a variable region of murine  
25 origin associated with a human constant region), humanized (a human immunoglobulin constant backbone together with hypervariable region of animal, e.g., murine, origin), and/or single chain. Both polyclonal and monospecific antibodies may also be in the form of immunoglobulin fragments, e.g., F(ab)'2  
30 or Fab fragments. The antibodies of the invention are of any isotype, e.g., IgG or IgA, and polyclonal antibodies are of a single isotype or a mixture of isotypes.

Antibodies against the polypeptides, homologs or fragments of the present invention are generated by immunization

of a mammal with a composition comprising said polypeptide, homolog or fragment. Such antibodies may be polyclonal or monoclonal. Methods to produce polyclonal or monoclonal antibodies are well known in the art. For a review, see

5 "Antibodies, A Laboratory Manual, Cold Spring Harbor Laboratory, Eds. E. Harlow and D. Lane (1988), and D.E. Yelton et al., 1981. Ann. Rev. Biochem. 50:657-680. For monoclonal antibodies, see Kohl and Milstein?...

The antibodies of the invention, which are raised to a

10 polypeptide or polypeptide derivative of the invention, are produced and identified using standard immunological assays, e.g., Western blot analysis, dot blot assay, or ELISA (see, e.g., Coligan et al., Current Protocols in Immunology (1994) John Wiley & Sons, Inc., New York, NY). The antibodies are used

15 in diagnostic methods to detect the presence of a *Chlamydia* antigen in a sample, such as a biological sample. The antibodies are also used in affinity chromatography for purifying a polypeptide or polypeptide derivative of the invention. As is discussed further below, such antibodies may

20 be used in prophylactic and therapeutic passive immunization methods.

Accordingly, a ninth aspect of the invention provides

(i) a reagent for detecting the presence of *Chlamydia* in a biological sample that contains an antibody, polypeptide, or

25 polypeptide derivative of the invention; and (ii) a diagnostic method for detecting the presence of *Chlamydia* in a biological sample, by contacting the biological sample with an antibody, a polypeptide, or a polypeptide derivative of the invention, such that an immune complex is formed, and by detecting such complex

30 to indicate the presence of *Chlamydia* in the sample or the organism from which the sample is derived.

Those skilled in the art will readily understand that the immune complex is formed between a component of the sample and the antibody, polypeptide, or polypeptide derivative, whichever

is used, and that any unbound material is removed prior to detecting the complex. It is understood that a polypeptide reagent is useful for detecting the presence of anti-*Chlamydia* antibodies in a sample, e.g., a blood sample, while an antibody of the invention is used for screening a sample, such as a gastric extract or biopsy, for the presence of *Chlamydia* polypeptides.

For diagnostic applications, the reagent (i.e., the antibody, polypeptide, or polypeptide derivative of the invention) is either in a free state or immobilized on a solid support, such as a tube, a bead, or any other conventional support used in the field. Immobilization is achieved using direct or indirect means. Direct means include passive adsorption (non-covalent binding) or covalent binding between the support and the reagent. By "indirect means" is meant that an anti-reagent compound that interacts with a reagent is first attached to the solid support. For example, if a polypeptide reagent is used, an antibody that binds to it can serve as an anti-reagent, provided that it binds to an epitope that is not involved in the recognition of antibodies in biological samples. Indirect means may also employ a ligand-receptor system, for example, where a molecule such as a vitamin is grafted onto the polypeptide reagent and the corresponding receptor immobilized on the solid phase. This is illustrated by the biotin-streptavidin system. Alternatively, a peptide tail is added chemically or by genetic engineering to the reagent and the grafted or fused product immobilized by passive adsorption or covalent linkage of the peptide tail.

Such diagnostic agents may be included in a kit which also comprises instructions for use. The reagent are labeled with a detection means which allows for the detection of the reagent when it is bound to its target. The detection means may be a fluorescent agent such as fluorescein isocyanate or fluorescein isothiocyanate, or an enzyme such as horse radish

peroxidase or luciferase or alkaline phosphatase, or a radioactive element such as  $^{125}\text{I}$  or  $^{51}\text{Cr}$ .

Accordingly, a tenth aspect of the invention provides a process for purifying, from a biological sample, a polypeptide or polypeptide derivative of the invention, which involves carrying out antibody-based affinity chromatography with the biological sample, wherein the antibody is a monospecific antibody of the invention.

For use in a purification process of the invention, the antibody is either polyclonal or monospecific, and preferably is of the IgG type. Purified IgGs is prepared from an antiserum using standard methods (see, e.g., Coligan *et al.*, *supra*). Conventional chromatography supports, as well as standard methods for grafting antibodies, are described in, e.g., Antibodies: A Laboratory Manual, D. Lane, E. Harlow, Eds. (1988) and outlined below.

Briefly, a biological sample, such as an *C. pneumoniae* extract preferably in a buffer solution, is applied to a chromatography material, preferably equilibrated with the buffer used to dilute the biological sample so that the polypeptide or polypeptide derivative of the invention (*i.e.*, the antigen) is allowed to adsorb onto the material. The chromatography material, such as a gel or a resin coupled to an antibody of the invention, is in either a batch form or a column. The unbound components are washed off and the antigen is then eluted with an appropriate elution buffer, such as a glycine buffer or a buffer containing a chaotropic agent, e.g., guanidine HCl, or high salt concentration (e.g., 3 M  $\text{MgCl}_2$ ). Eluted fractions are recovered and the presence of the antigen is detected, e.g., by measuring the absorbance at 280 nm.

An eleventh aspect of the invention provides (i) a composition of matter comprising a monospecific antibody of the invention, together with a diluent or carrier; (ii) a pharmaceutical composition comprising a therapeutically or



prophylactically effective amount of a monospecific antibody of the invention, and (iii) a method for treating or preventing a *Chlamydia* (e.g., *C. trachomatis*, *C. psittaci*, *C. pneumoniae* or *C. pecorum*) infection, by administering a therapeutic or  
5 prophylactic amount of a monospecific antibody of the invention to an infected individual. Additionally, the eleventh aspect of the invention encompasses the use of a monospecific antibody of the invention in the preparation of a medicament for treating or preventing *Chlamydia* infection.

10 The monospecific antibody is either polyclonal or monoclonal, preferably of the IgA isotype (predominantly). In passive immunization, the antibody is administered to a mucosal surface of a mammal, e.g., the gastric mucosa, e.g., orally or intragastrically, advantageously, in the presence of a  
15 bicarbonate buffer. Alternatively, systemic administration, not requiring a bicarbonate buffer, is carried out. A monospecific antibody of the invention is administered as a single active component or as a mixture with at least one monospecific antibody specific for a different *Chlamydia* polypeptide. The  
20 amount of antibody and the particular regimen used are readily determined by one skilled in the art. For example, daily administration of about 100 to 1,000 mg of antibodies over one week, or three doses per day of about 100 to 1,000 mg of antibodies over two or three days, are effective regimens for  
25 most purposes.

Therapeutic or prophylactic efficacy are evaluated using standard methods in the art, e.g., by measuring induction of a mucosal immune response or induction of protective and/or therapeutic immunity, using, e.g., the *C. pneumoniae* mouse  
30 model. Those skilled in the art will readily recognize that the *C. pneumoniae* strain of the model may be replaced with another *Chlamydia* strain. For example, the efficacy of DNA molecules and polypeptides from *C. pneumoniae* is preferably evaluated in a mouse model using *C. pneumoniae* strain. Protection is

determined by comparing the degree of *Chlamydia* infection to that of a control group. Protection is shown when infection is reduced by comparison to the control group. Such an evaluation is made for polynucleotides, vaccine vectors, polypeptides and 5 derivatives thereof, as well as antibodies of the invention.

Adjuvants useful in any of the vaccine compositions described above are as follows.

Adjuvants for parenteral administration include aluminum compounds, such as aluminum hydroxide, aluminum phosphate, and 10 aluminum hydroxy phosphate. The antigen is precipitated with, or adsorbed onto, the aluminum compound according to standard protocols. Other adjuvants, such as RIBI (ImmunoChem, Hamilton, MT), is used in parenteral administration.

Adjuvants for mucosal administration include bacterial 15 toxins, e.g., the cholera toxin (CT), the *E. coli* heat-labile toxin (LT), the *Clostridium difficile* toxin A and the pertussis toxin (PT), or combinations, subunits, toxoids, or mutants thereof such as a purified preparation of native cholera toxin subunit B (CTB). Fragments, homologs, derivatives, and fusions 20 to any of these toxins are also suitable, provided that they retain adjuvant activity. Preferably, a mutant having reduced toxicity is used. Suitable mutants are described, e.g., in WO 95/17211 (Arg-7-Lys CT mutant), WO 96/6627 (Arg-192-Gly LT mutant), and WO 95/34323 (Arg-9-Lys and Glu-129-Gly PT mutant). 25 Additional LT mutants that are used in the methods and compositions of the invention include, e.g., Ser-63-Lys, Ala-69-Gly, Glu-110-Asp, and Glu-112-Asp mutants. Other adjuvants, such as a bacterial monophosphoryl lipid A (MPLA) of, e.g., *E. coli*, *Salmonella minnesota*, *Salmonella typhimurium*, or *Shigella* 30 *flexneri*; saponins, or polylactide glycolide (PLGA) microspheres, is also be used in mucosal administration.

Adjuvants useful for both mucosal and parenteral administrations include polyphosphazene (WO 95/2415), DC-chol (3

b-(N-(N',N'-dimethyl aminomethane)-carbamoyl) cholesterol; U.S. Patent No. 5,283,185 and WO 96/14831) and QS-21 (WO 88/9336).

Any pharmaceutical composition of the invention containing a polynucleotide, a polypeptide, a polypeptide  
5 derivative, or an antibody of the invention, is manufactured in a conventional manner. In particular, it is formulated with a pharmaceutically acceptable diluent or carrier, e.g., water or a saline solution such as phosphate buffer saline. In general, a diluent or carrier is selected on the basis of the mode and  
10 route of administration, and standard pharmaceutical practice. Suitable pharmaceutical carriers or diluents, as well as pharmaceutical necessities for their use in pharmaceutical formulations, are described in *Remington's Pharmaceutical Sciences*, a standard reference text in this field and in the  
15 USP/NF.

The invention also includes methods in which *Chlamydia* infection are treated by oral administration of a *Chlamydia* polypeptide of the invention and a mucosal adjuvant, in combination with an antibiotic, an antacid, sucralfate, or a  
20 combination thereof. Examples of such compounds that can be administered with the vaccine antigen and the adjuvant are antibiotics, including, e.g., macrolides, tetracyclines, and derivatives thereof (specific examples of antibiotics that can be used include azithromycin or doxycyclin or immunomodulators  
25 such as cytokines or steroids). In addition, compounds containing more than one of the above-listed components coupled together, are used. The invention also includes compositions for carrying out these methods, i.e., compositions containing a *Chlamydia* antigen (or antigens) of the invention, an adjuvant,  
30 and one or more of the above-listed compounds, in a pharmaceutically acceptable carrier or diluent.

Amounts of the above-listed compounds used in the methods and compositions of the invention are readily determined by one skilled in the art. Treatment/immunization schedules are also

known and readily designed by one skilled in the art. For example, the non-vaccine components can be administered on days 1-14, and the vaccine antigen + adjuvant can be administered on days 7, 14, 21, and 28.

**REFERENCES**

1. Grayston et al. (1995) Journal of Infectious Diseases 168:1231
2. Campos et al. (1995) Investigation of Ophthalmology and  
5 Visual Science 36:1477
3. Grayston et al (1990) Journal of Infectious Diseases 161:618
4. Marrie (1993) Clinical Infectious Diseases. 18:501
5. Wang et al (1986) Chlamydial infections. Cambridge  
University Press, Cambridge. p. 329
- 10 6. Saikku et al.(1988) Lancet;ii:983
7. Thom et al. (1992) JAMA 268:68
8. Linnanmaki et al. (1993), Circulation 87:1030
9. Saikku et al. (1992)Annals Internal Medicine 116:273
10. Melnick et al(1993) American Journal of Medicine 95:499 11
- 15 Shor et al. (1992) South African. Medical Journal 82:158
12. Kuo et al. (1993) Journal of Infectious Diseases 167:841
13. Kuo et al. (1993) Arteriosclerosis and Thrombosis 13:1500
14. Campbell et al (1995) Journal of Infectious Diseases 172:585
15. Chiu et al. Circulation, 1997 (In Press).
- 20 16. Ramirez et al (1996) Annals of Internal Medicine 125:979
17. Jackson et al. Abst. K121, p272, 36<sup>th</sup> ICAAC, 15-18 Sept.  
1996, New Orleans.
18. Fong et al (1997) Journal of Clinical Microbiology 35:48
19. Hahn DL, et al. Evidence for Chlamydia pneumoniae infection  
25 in steroid-dependent asthma. Ann Allergy Asthma Immunol. 1998  
Jan; 80(1): 45-49.
20. Hahn DL, et al. Association of Chlamydia pneumoniae IgA  
antibodies with recently symptomatic asthma. Epidemiol Infect.  
1996 Dec; 117(3): 513-517.
- 30 21. Bjornsson E, et al. Serology of chlamydia in relation to  
asthma and bronchial hyperresponsiveness. Scand J Infect Dis.  
1996; 28(1): 63-69.

22. Hahn DL. Treatment of *Chlamydia pneumoniae* infection in adult asthma: a before-after trial. J Fam Pract. 1995 Oct; 41(4): 345-351.
23. Allegra L, et al. Acute exacerbations of asthma in adults: role of *Chlamydia pneumoniae* infection. Eur Respir J. 1994 Dec; 7(12): 2165-2168.
24. Hahn DL, et al. Association of *Chlamydia pneumoniae* (strain TWAR) infection with wheezing, asthmatic bronchitis, and adult-onset asthma. JAMA. 1991 Jul 10; 266(2): 225-230.
- 10 25. Pal et al. (1996) Infection and Immunity. 64:5341
26. Jones et al. (1995) Vaccine 13:715
27. Igietsemes et al. (1993) Immunology 5:317
28. Igietseme et al (1993) Regional Immunology 5:317
29. Magee et al (1993) Regional Immunology 5: 305
- 15 30. Landers et al (1991) Infection & Immunity 59:3774
31. Magee et al (1995) Infection & Immunity 63:516
32. Cotter et al. (1995) Infection and Immunity 63:4704
33. Campbell et al (1990) Infection and Immunity 58:93
34. McCafferty et al (1995) Infection and Immunity 63:2387-9.
- 20 35. Knudsen et al (1996) Third Meeting of the European Society for *Chlamydia* Research, Vienna
36. Wiedmann-Al-Ahmad M, et al. Reactions of polyclonal and neutralizing anti-p54 monoclonal antibodies with an isolated, species-specific 54-kilodalton protein of *Chlamydia pneumoniae*.
- 25 Clin Diagn Lab Immunol. 1997 Nov; 4(6): 700-704
37. Hughes et al., 1992. Infect. Immun. 60(9):3497,
38. Dion et al., 1990. Virology 179:474-477,
39. Snijders et al., 1991. J. Gen. Virol. 72:557-565,
40. Langeveld et al., Vaccine 12(15):1473-1480, 1994
- 30 41. Ausubel et al., Current Protocols in Molecular Biology, John Wiley & Sons Inc., 1994
42. Kunkel et al. Proc. Natl. Acad. Sci. USA (1985) 82:448
43. Silhavy et al. Experiments with Gene Fusions, Cold Spring Harbor Laboratory Press, 1984

44. Davis et al. A Manual for Genetic Engineering: Advanced  
Bacterial Genetics, Cold Spring Harbor Laboratory Press, 1980)
45. Casey & Davidson, Nucl. Acid Res. (1977) 4:1539
46. Cagnon et al., Protein Engineering (1991) 4(7):843
- 5 47. Takase et al., J. Bact. (1987) 169:5692

CLAIMS

1. A nucleic acid molecule comprising a nucleic acid  
sequence which encodes a polypeptide selected from any of:
- 5 (a) SEQ ID Nos: 27 to 45;
- (b) an immunogenic fragment comprising at least 12  
consecutive amino acids from a polypeptide of (a); and
- (c) a polypeptide of (a) or (b) which has been modified to  
improve its immunogenicity, wherein said modified  
10 polypeptide is at least 75% identical in amino acid  
sequence to the corresponding polypeptide of (a) or  
(b).
2. A nucleic acid molecule comprising a nucleic acid  
15 sequence selected from any of:
- (a) SEQ ID Nos: 1 to 26;
- (b) a sequence which encodes a polypeptide encoded by any  
one of SEQ ID Nos: 1 to 26;
- (c) a sequence comprising at least 38 consecutive  
20 nucleotides from any one of the nucleic acid sequences  
of (a) and (b); and
- (d) a sequence which encodes a polypeptide which is at  
least 75% identical in amino acid sequence to any one  
of the polypeptides encoded by SEQ ID Nos: 1 to 26.

25



3. A nucleic acid molecule comprising a nucleic acid sequence which encodes a fusion protein, said fusion protein comprising a polypeptide encoded by a nucleic acid molecule according to claim 1 and an additional  
5 polypeptide.
4. A nucleic acid molecule according to claim 1, operatively linked to one or more expression control sequences.  
10
5. A vaccine comprising at least one first nucleic acid according to any one of claims 1 to 4 and a vaccine vector wherein each first nucleic acid is expressed as a polypeptide, the vaccine optionally comprising a second  
15 nucleic acid encoding an additional polypeptide which enhances the immune response to the polypeptide expressed by said first nucleic acid.
6. The vaccine of claim 5 wherein the second nucleic acid  
20 encodes an additional *Chlamydia* polypeptide.
7. A pharmaceutical composition comprising a nucleic acid according to any one of claims 1 to 5 and a pharmaceutically acceptable carrier.  
25

8. A pharmaceutical composition comprising a vaccine according to claim 5 or 6 and a pharmaceutically acceptable carrier.
- 5 9. A unicellular host transformed with the nucleic acid molecule of claim 4.
10. A nucleic acid probe of 5 to 100 nucleotides which hybridizes under stringent conditions to any one of nucleic acid molecules of SEQ ID Nos: 1 to 26, or to a homolog or complementary or anti-sense sequence of said nucleic acid molecule.
- 10 11. A primer of 10 to 40 nucleotides which hybridizes under stringent conditions to any one of nucleic acid molecules of SEQ ID Nos: 1 to 26, or to a homolog or complementary or anti-sense sequence of said nucleic acid molecule.
- 15 12. A polypeptide encoded by a nucleic acid sequence according to any one of claims 1 to 4.
- 20 13. A polypeptide comprising an amino acid sequence selected from any of:
- 25 (a) SEQ ID Nos: 27 to 45;

- (b) an immunogenic fragment comprising at least 12 consecutive amino acids from a polypeptide of (a); and
- (c) a polypeptide of (a) or (b) which has been modified to improve its immunogenicity, wherein said modified polypeptide is at least 75% identical in amino acid sequence to the corresponding polypeptide of (a) or (b).
14. A fusion polypeptide comprising a polypeptide of claim 12 or 13 and an additional polypeptide.
15. A method for producing a polypeptide of claim 12 or 13, comprising the step of culturing a unicellular host according to claim 9.
16. An antibody against the polypeptide of any one of claims 12 to 14.
17. A vaccine comprising at least one first polypeptide according to any one of claims 12 to 14 and a pharmaceutically acceptable carrier, optionally comprising a second polypeptide which enhances the immune response to the first polypeptide.
18. The vaccine of claim 17 wherein the second polypeptide comprises an additional *Chlamydia* polypeptide.

19. A pharmaceutical composition comprising a polypeptide according to any one of claims 12 to 14 and a pharmaceutically acceptable carrier.

5

20. A pharmaceutical composition comprising a vaccine according to claim 17 or 18 and a pharmaceutically acceptable carrier.

- 10 21. A pharmaceutical composition comprising an antibody according to claim 16 and a pharmaceutically acceptable carrier.

22. A method for preventing or treating *Chlamydia* infection using:
- 15 (a) the nucleic acid of any one of claims 1 to 4;
- (b) the vaccine of any one of claims 5, 6, 17 and 18;
- (c) the pharmaceutical composition of any one of claims 7, 8, 19 to 21;
- 20 (d) the polypeptide of any one of claims 12 to 14; or
- (e) the antibody of claim 16.

23. A method of detecting *Chlamydia* infection comprising the step of assaying a body fluid of a mammal to be tested, with a component selected from any one of:
- 25 (a) the nucleic acid of any one of claims 1 to 4;

- (b) the polypeptide of any one of claims 12 to 14; and
- (c) the antibody of claim 16.

24. A diagnostic kit comprising instructions for use and a  
5 component selected from any one of:

- (a) the nucleic acid of any one of claims 1 to 4;
- (b) the polypeptide of any one of claims 12 to 14; and  
the antibody of claim 16.

Figure 1: CPN100397

```

attttaacgt gcgtatcatt tgtgactaag agatagactt gctttcttta tctatcttct 60
gtattggaaa gaaagcccct tgagggaaaa aaaggttggt atg aag att cca ctc 115
                                         Met Lys Ile Pro Leu
                                         1                               5,

cgc ttt tta ttg ata tca tta gta cct acg ctt tct atg tcg aat tta 163
Arg Phe Leu Leu Ile Ser Leu Val Pro Thr Leu Ser Met Ser Asn Leu
                               10                               15                               20

tta gga gct gct act acc gaa gag tta tcg gct agc aat agc ttc gat 211
Leu Gly Ala Ala Thr Thr Glu Glu Leu Ser Ala Ser Asn Ser Phe Asp
                               25                               30                               35

gga act aca tca aca aca agc ttt tct agt aaa aca tca tcg gct aca 259
Gly Thr Thr Ser Thr Thr Ser Phe Ser Ser Lys Thr Ser Ser Ala Thr
                               40                               45                               50

gat ggc acc aat tat gtt ttt aaa gat tct gta gtt ata gaa aat gta 307
Asp Gly Thr Asn Tyr Val Phe Lys Asp Ser Val Val Ile Glu Asn Val
                               55                               60                               65

ccc aaa aca ggg gaa act cag tct act agt tgt ttt aaa aat gac gct 355
Pro Lys Thr Gly Glu Thr Gln Ser Thr Ser Cys Phe Lys Asn Asp Ala
                               70                               75                               80                               85

gca gct gga gat cta aat ttc tta gga ggg gga ttt tct ttc aca ttt 403
Ala Ala Gly Asp Leu Asn Phe Leu Gly Gly Gly Phe Ser Phe Thr Phe
                               90                               95                               100

agc aat atc gat gca acc acg gct tct gga gct gct att gga agt gaa 451
Ser Asn Ile Asp Ala Thr Thr Ala Ser Gly Ala Ala Ile Gly Ser Glu
                               105                               110                               115

gca gct aat aag aca gtc acg tta tca gga ttt tcg gca ctt tct ttt 499
Ala Ala Asn Lys Thr Val Thr Leu Ser Gly Phe Ser Ala Leu Ser Phe
                               120                               125                               130

ctt aaa tcc cca gca agt aca gtg act aat gga ttg gga gct atc aat 547
Leu Lys Ser Pro Ala Ser Thr Val Thr Asn Gly Leu Gly Ala Ile Asn
                               135                               140                               145

gtt aaa ggg aat tta agc cta ttg gat aat gat aag gta ttg att cag 595
Val Lys Gly Asn Leu Ser Leu Leu Asp Asn Asp Lys Val Leu Ile Gln
                               150                               155                               160                               165

gac aat ttc tca aca gga gat ggc gga gca att aat tgt gca ggc tcc 643
Asp Asn Phe Ser Thr Gly Asp Gly Gly Ala Ile Asn Cys Ala Gly Ser
                               170                               175                               180

ttg aag atc gca aac aat aag tcc ctt tct ttt att gga aat agt tct 691
Leu Lys Ile Ala Asn Asn Lys Ser Leu Ser Phe Ile Gly Asn Ser Ser
                               185                               190                               195

```

Fig. 1 (con't)

tca aca cgt ggc gga gcg att cat acc aaa aac ctc aca cta tct tct	739
Ser Thr Arg Gly Gly Ala Ile His Thr Lys Asn Leu Thr Leu Ser Ser	
200 205 210	
ggg ggg gaa act cta ttt cag ggg aat aca gcg cct acg gct gct ggt	787
Gly Gly Glu Thr Leu Phe Gln Gly Asn Thr Ala Pro Thr Ala Ala Gly	
215 220 225	
aaa gga ggt gct atc gcg att gca gac tct ggc acc cta tcc att tct	835
Lys Gly Gly Ala Ile Ala Ile Ala Asp Ser Gly Thr Leu Ser Ile Ser	
230 235 240 245	
gga gac agt ggc gac att atc ttt gaa ggc aat acg ata gga gct aca	883
Gly Asp Ser Gly Asp Ile Ile Phe Glu Gly Asn Thr Ile Gly Ala Thr	
250 255 260	
gga acc gtc tct cat agt gct att gat tta gga act agc gct aag ata	931
Gly Thr Val Ser His Ser Ala Ile Asp Leu Gly Thr Ser Ala Lys Ile	
265 270 275	
act gcg tta cgt gct gcg caa gga cat acg ata tac ttt tat gat ccg	979
Thr Ala Leu Arg Ala Ala Gln Gly His Thr Ile Tyr Phe Tyr Asp Pro	
280 285 290	
att act gta aca gga tcg aca tct gtt gct gat gct ctc aat att aat	1027
Ile Thr Val Thr Gly Ser Thr Ser Val Ala Asp Ala Leu Asn Ile Asn	
295 300 305	
agc cct gat act gga gat aac aaa gag tat acg gga acc ata gtc ttt	1075
Ser Pro Asp Thr Gly Asp Asn Lys Glu Tyr Thr Gly Thr Ile Val Phe	
310 315 320 325	
tct gga gag aag ctc acg gag gca gaa gct aaa gat gag aag aac cgc	1123
Ser Gly Glu Lys Leu Thr Glu Ala Glu Ala Lys Asp Glu Lys Asn Arg	
330 335 340	
act tct aaa tta ctt caa aat gtt gct ttt aaa aat ggg act gta gtt	1171
Thr Ser Lys Leu Leu Gln Asn Val Ala Phe Lys Asn Gly Thr Val Val	
345 350 355	
tta aaa ggt gat gtc gtt tta agt gcg aac ggt ttc tct cag gat gca	1219
Leu Lys Gly Asp Val Val Leu Ser Ala Asn Gly Phe Ser Gln Asp Ala	
360 365 370	
aac tct aag ttg att atg gat tta ggg acg tcg ttg gtt gca aac acc	1267
Asn Ser Lys Leu Ile Met Asp Leu Gly Thr Ser Leu Val Ala Asn Thr	
375 380 385	
gaa agt atc gag tta acg aat ttg gaa att aat ata gac tct ctc agg	1315
Glu Ser Ile Glu Leu Thr Asn Leu Glu Ile Asn Ile Asp Ser Leu Arg	
390 395 400 405	

Fig. 1 (con't)

aac ggg aaa aag ata aaa ctc agt gct gcc aca gct cag aaa gat att	1363
Asn Gly Lys Lys Ile Lys Leu Ser Ala Ala Thr Ala Gln Lys Asp Ile	
410 415 420	
cgt ata gat cgt cct gtt gta ctg gca att agc gat gag agt ttt tat	1411
Arg Ile Asp Arg Pro Val Val Leu Ala Ile Ser Asp Glu Ser Phe Tyr	
425 430 435	
caa aat ggc ttt ttg aat gag gac cat tcc tat gat ggg att ctt gag	1459
Gln Asn Gly Phe Leu Asn Glu Asp His Ser Tyr Asp Gly Ile Leu Glu	
440 445 450	
tta gat gct ggg aaa gac atc gtg att tct gca gat tct cgc agt ata	1507
Leu Asp Ala Gly Lys Asp Ile Val Ile Ser Ala Asp Ser Arg Ser Ile	
455 460 465	
gat gct gta caa tct ccg tat ggc tat cag gga aag tgg acg atc aat	1555
Asp Ala Val Gln Ser Pro Tyr Gly Tyr Gln Gly Lys Trp Thr Ile Asn	
470 475 480 485	
tgg tct act gat gat aag aaa gct acg gtt tct tgg gcg aag cag agt	1603
Trp Ser Thr Asp Asp Lys Lys Ala Thr Val Ser Trp Ala Lys Gln Ser	
490 495 500	
ttt aat ccc act gct gag cag gag gct ccg tta gtt cct aat ctt ctt	1651
Phe Asn Pro Thr Ala Glu Gln Glu Ala Pro Leu Val Pro Asn Leu Leu	
505 510 515	
tgg ggt tct ttt ata gat gtt cgt tcc ttc cag aat ttt ata gag cta	1699
Trp Gly Ser Phe Ile Asp Val Arg Ser Phe Gln Asn Phe Ile Glu Leu	
520 525 530	
ggg act gaa ggt gct cct tac gaa aag aga ttt tgg gtt gca ggc att	1747
Gly Thr Glu Gly Ala Pro Tyr Glu Lys Arg Phe Trp Val Ala Gly Ile	
535 540 545	
tcc aat gtt ttg cat agg agc ggt cgt gaa aat caa agg aaa ttc cgt	1795
Ser Asn Val Leu His Arg Ser Gly Arg Glu Asn Gln Arg Lys Phe Arg	
550 555 560 565	
cat gtg agt gga ggt gct gta gta ggt gct agc acg agg atg ccg ggt	1843
His Val Ser Gly Gly Ala Val Val Gly Ala Ser Thr Arg Met Pro Gly	
570 575 580	
ggg gat acc ttg tct ctg ggt ttt gct cag ctc ttt gcg cgt gac aaa	1891
Gly Asp Thr Leu Ser Leu Gly Phe Ala Gln Leu Phe Ala Arg Asp Lys	
585 590 595	
gac tac ttt atg aat acc aat ttc gca aag acc tac gca gga tct tta	1939
Asp Tyr Phe Met Asn Thr Asn Phe Ala Lys Thr Tyr Ala Gly Ser Leu	
600 605 610	



Fig. 1 (con't)

cgt ttg cag cac gat gct tcc cta tac tct gtg gtg agt atc ctt tta	1987
Arg Leu Gln His Asp Ala Ser Leu Tyr Ser Val Val Ser Ile Leu Leu	
615 620 625	
gga gag gga gga ctc cgc gag atc ctg ttg cct tat gtt tcc aag act	2035
Gly Glu Gly Gly Leu Arg Glu Ile Leu Leu Pro Tyr Val Ser Lys Thr	
630 635 640 645	
ctg ccg tgc tct ttc tat ggg cag ctt agc tac ggc cat acg gat cat	2083
Leu Pro Cys Ser Phe Tyr Gly Gln Leu Ser Tyr Gly His Thr Asp His	
650 655 660	
cgc atg aag acc gag tct cta ccc ccc ccc ccc ccg acg ctc tcg acg	2131
Arg Met Lys Thr Glu Ser Leu Pro Pro Pro Pro Pro Thr Leu Ser Thr	
665 670 675	
gat cat act tct tgg gga gga tat gtc tgg gct gga gag ctg gga act	2179
Asp His Thr Ser Trp Gly Gly Tyr Val Trp Ala Gly Glu Leu Gly Thr	
680 685 690	
cga gtt gct gtt gaa aat acc agc ggc aga gga ttt ttc caa gag tac	2227
Arg Val Ala Val Glu Asn Thr Ser Gly Arg Gly Phe Phe Gln Glu Tyr	
695 700 705	
act cca ttt gta aaa gtc caa gct gtt tac gct cgc caa gat agc ttt	2275
Thr Pro Phe Val Lys Val Gln Ala Val Tyr Ala Arg Gln Asp Ser Phe	
710 715 720 725	
gta gaa cta gga gct atc agt cgt gat ttt agt gat tcg cat ctt tat	2323
Val Glu Leu Gly Ala Ile Ser Arg Asp Phe Ser Asp Ser His Leu Tyr	
730 735 740	
aac ctt gcg att cct ctt gga atc aag tta gag aaa cgg ttt gca gag	2371
Asn Leu Ala Ile Pro Leu Gly Ile Lys Leu Glu Lys Arg Phe Ala Glu	
745 750 755	
caa tat tat cat gtt gta gcg atg tat tct cca gat gtt tgt cgt agt	2419
Gln Tyr Tyr His Val Val Ala Met Tyr Ser Pro Asp Val Cys Arg Ser	
760 765 770	
aac ccc aaa tgt acg act acc cta ctt tcc aac caa ggg agt tgg aag	2467
Asn Pro Lys Cys Thr Thr Thr Leu Leu Ser Asn Gln Gly Ser Trp Lys	
775 780 785	
acc aaa ggt tcg aac tta gca aga cag gct ggt att gtt cag gcc tca	2515
Thr Lys Gly Ser Asn Leu Ala Arg Gln Ala Gly Ile Val Gln Ala Ser	
790 795 800 805	
ggg ttt cga tct ttg gga gct gca gca gag ctt ttc ggg aac ttt ggc	2563
Gly Phe Arg Ser Leu Gly Ala Ala Ala Glu Leu Phe Gly Asn Phe Gly	
810 815 820	

Fig. 1 (con't)

```

ttt gaa tgg cgg gga tct tct cgt agc tat aat gta gat gcg ggt agc 2611
Phe Glu Trp Arg Gly Ser Ser Arg Ser Tyr Asn Val Asp Ala Gly Ser
      825                      830                      835

aaa atc aaa ttt tagcgatttc tctttcgatg ctatTTTTcc atggctattt 2663
Lys Ile Lys Phe
      840

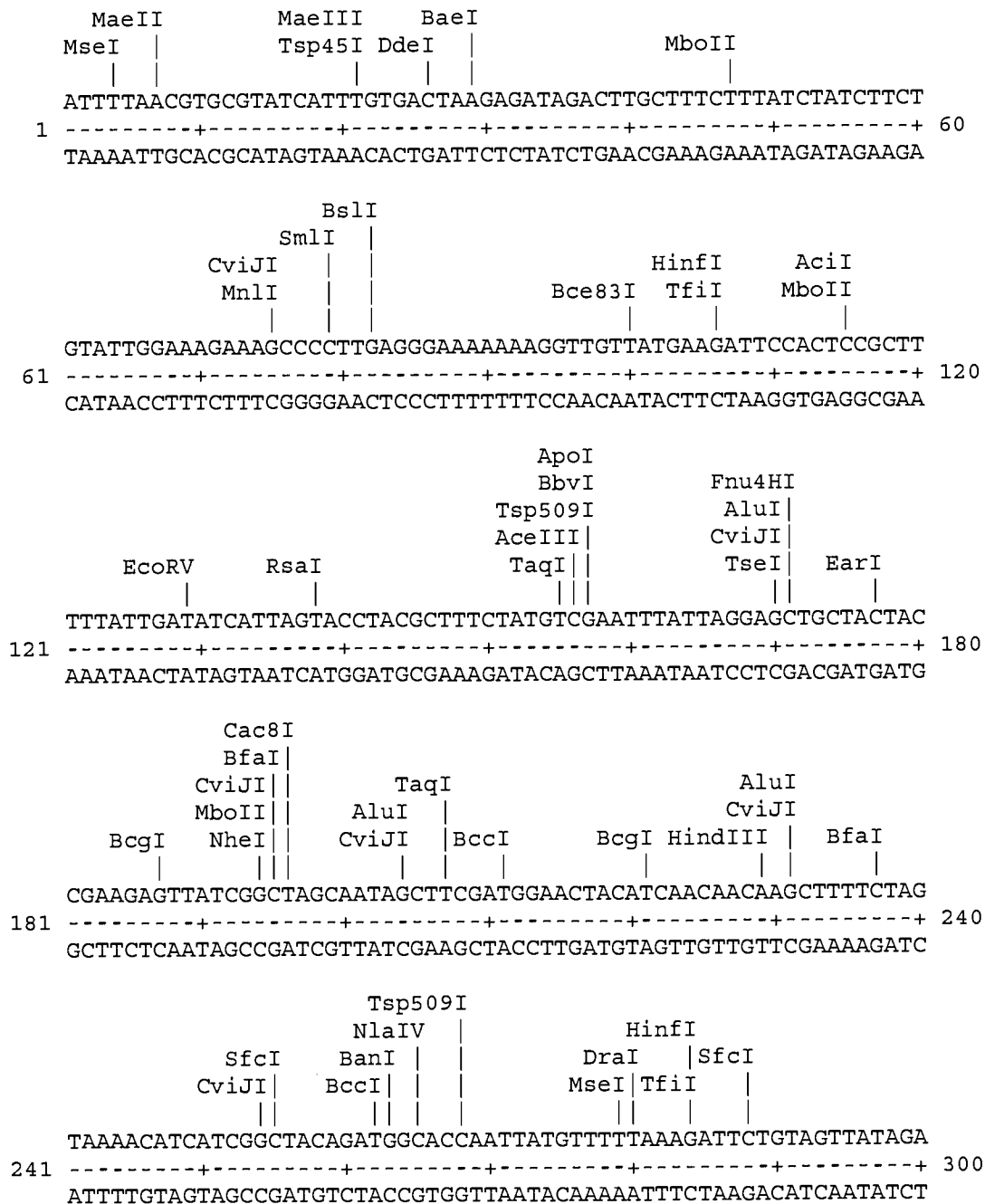
ttaaaatgat agccatgggtt atagatacgt agtccttatt tcaaagaaga cactgttgca 2723

ttagatacgc tctctgatcc ctcaaaa 2750

```

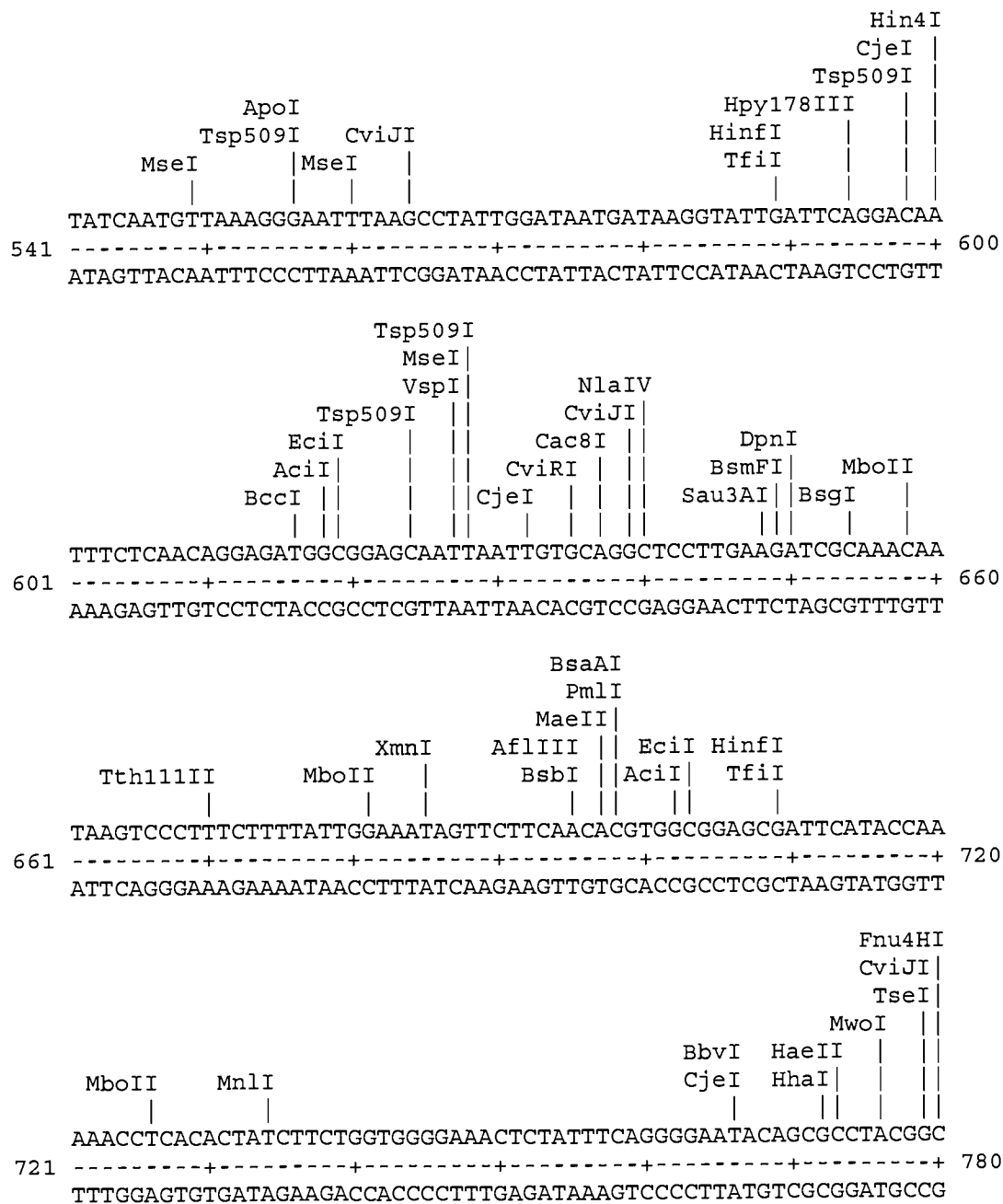
Figure 2 (RY-32)

## Restriction Enzyme analysis of CPN100397



AluI  
CviJI  
HgaI  
MspAII  
PvuII  
PstI  
DraI  
MseI  
BbvI  
BseMII  
BfaI  
SpeI  
RsaI  
BslI  
DdeI  
AccI  
Fnu4HI  
CviRI  
TseI  
Fnu4HI  
SfcI  
TseI  
ApoI  
Tsp509I  
BbvI  
DpnI  
BglII  
BstYI  
Sau3AI  
DdeI  
MnlI  
BpmI  
SfaNI  
AceIII  
BsaJI  
BstDSI  
CviRI  
CjeI  
ClaI  
TaqI  
TGGAGATCTAAATTTCTTAGGAGGGGGATTTTCTTTACATTTAGCAATATCGATGCAAC  
ACCTCTAGATTTAAAGAATCCTCCCCCTAAAGAAAGTGTAATCGTTATAGCTACGTTG  
Bcefi  
Fnu4HI  
AluI  
CviJI  
TseI  
MwoI  
Hpy178III  
CviJI  
MwoI  
BbvI  
AluI  
CviJI  
Fnu4HI  
TseI  
BpmI  
CjeI  
MaeIII  
TaaI  
Tsp45I  
BbvI  
Hpy178III  
MaeII  
CACGGCTTCTGGAGCTGCTATTGGAAGTGAAGCAGCTAATAAGACAGTCACGTTATCAGG  
GTGCCGAAGACCTCGACGATAACCTTCACTTCGTCGATTATTCTGTCTAGTGAATAGTCC  
MaeIII  
TaaI  
Tsp45I  
RsaI  
TatI  
MseI  
TspRI  
AluI  
CviJI  
ATTTTCGGCACTTTCTTTTCTTAAATCCCCAGCAAGTACAGTACTAATGGATTGGGAGC  
TAAAGCCGTGAAAGAAAAGAATTTAGGGTTCGTTTCATGTCAGTACTGATTACCTAACCTCG

Fig. 2 (con't)



[illegible]

Fig. 2 (con't)

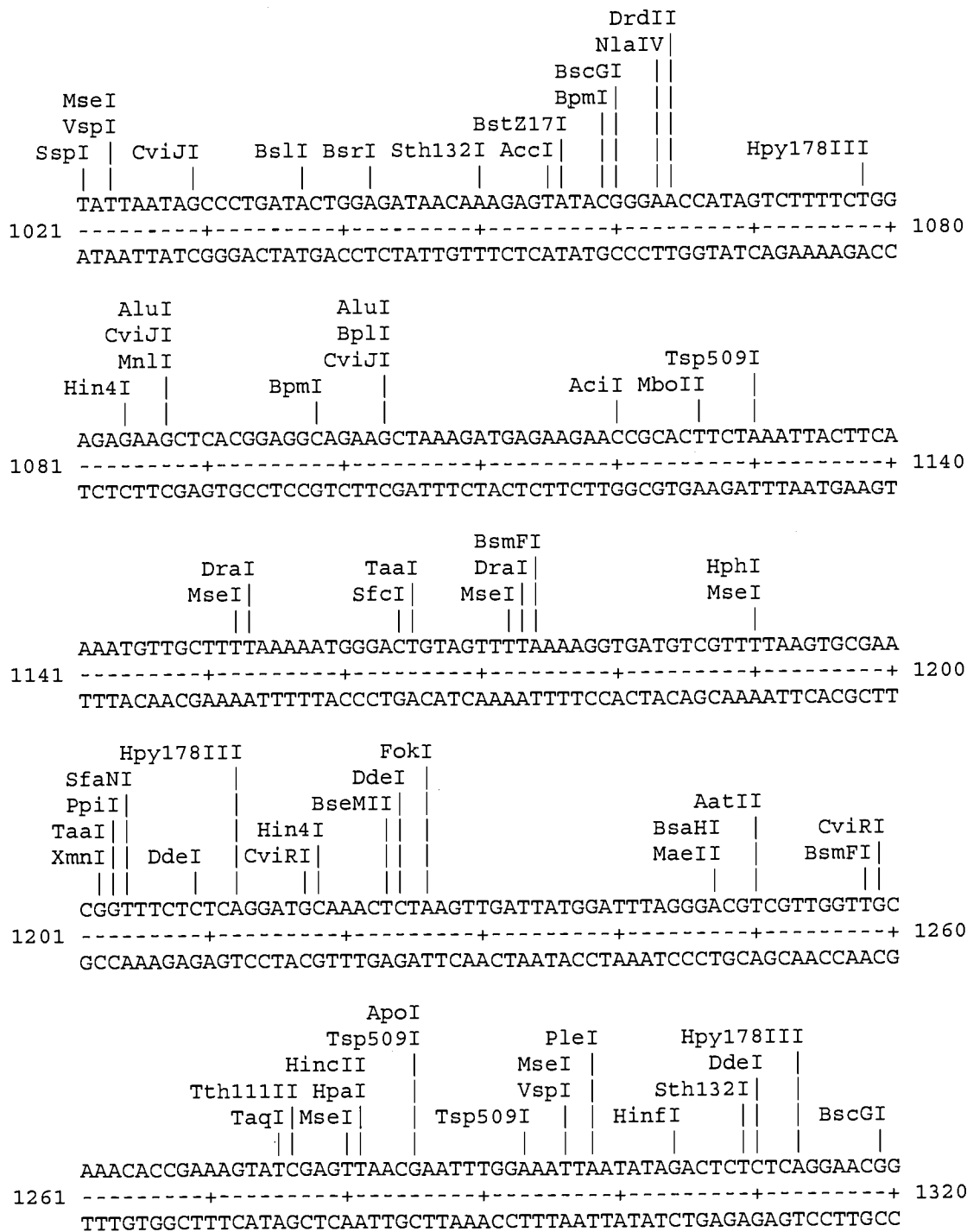


Fig. 2 (con't)

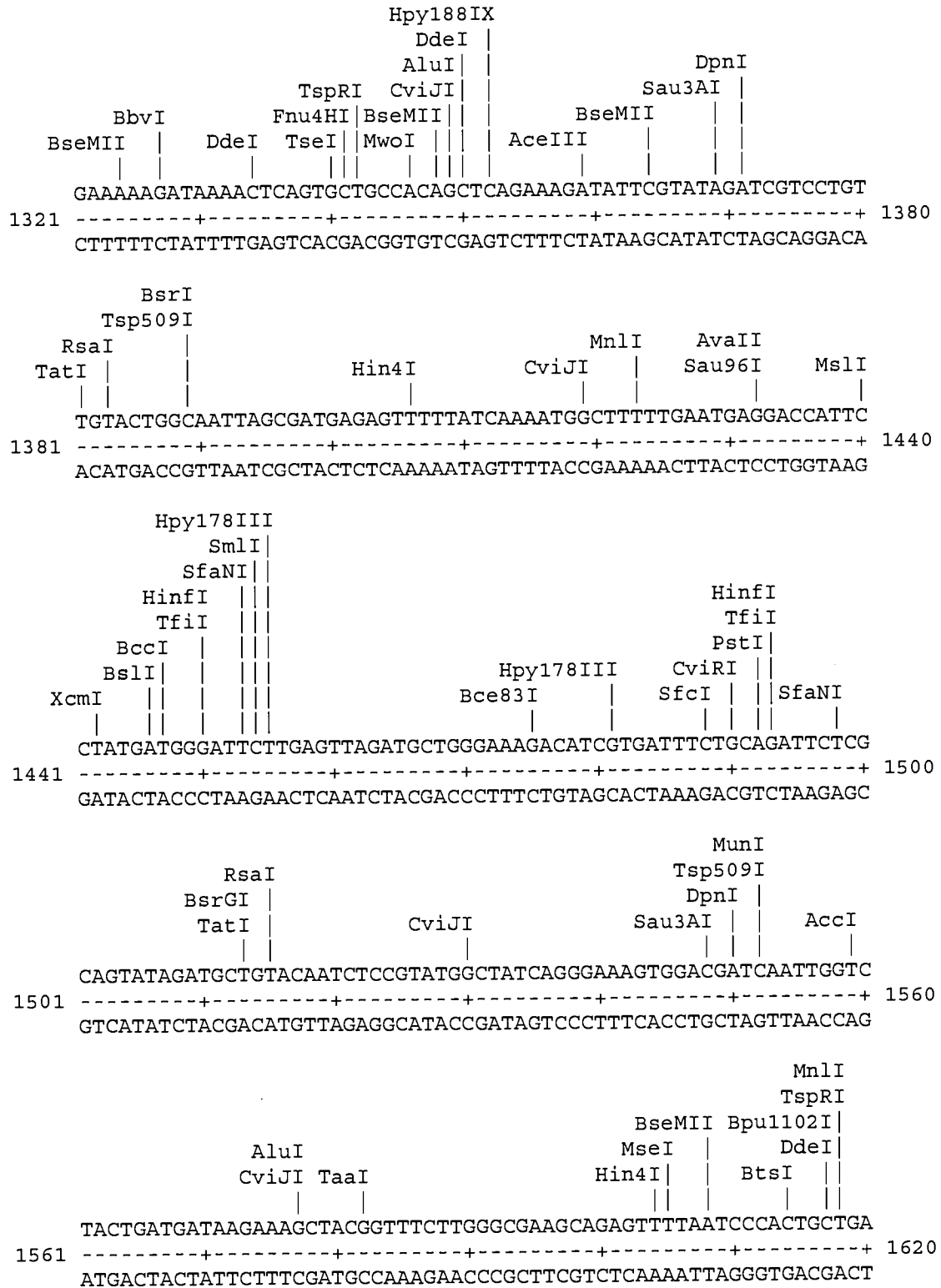




Fig. 2 (con't)

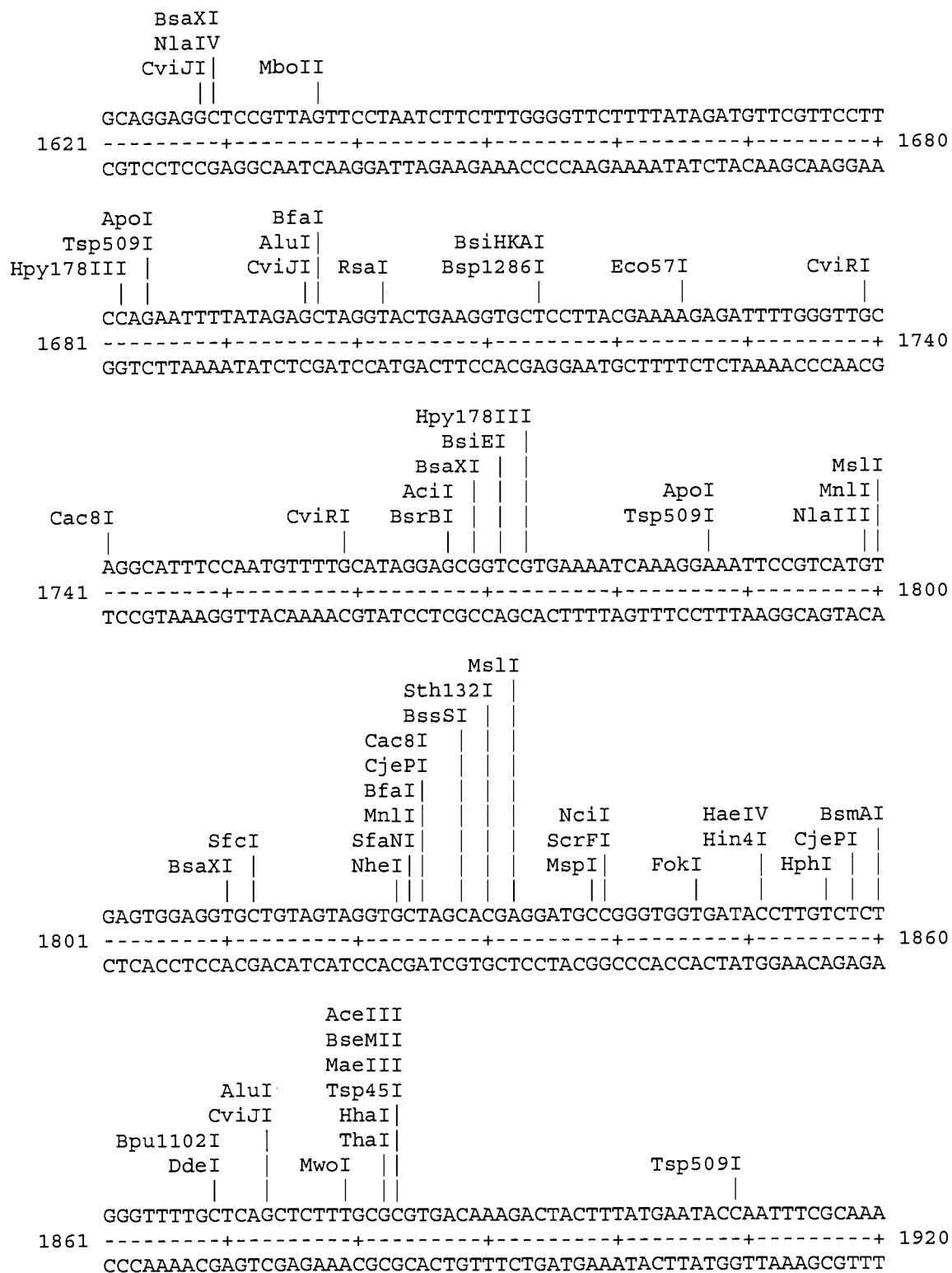
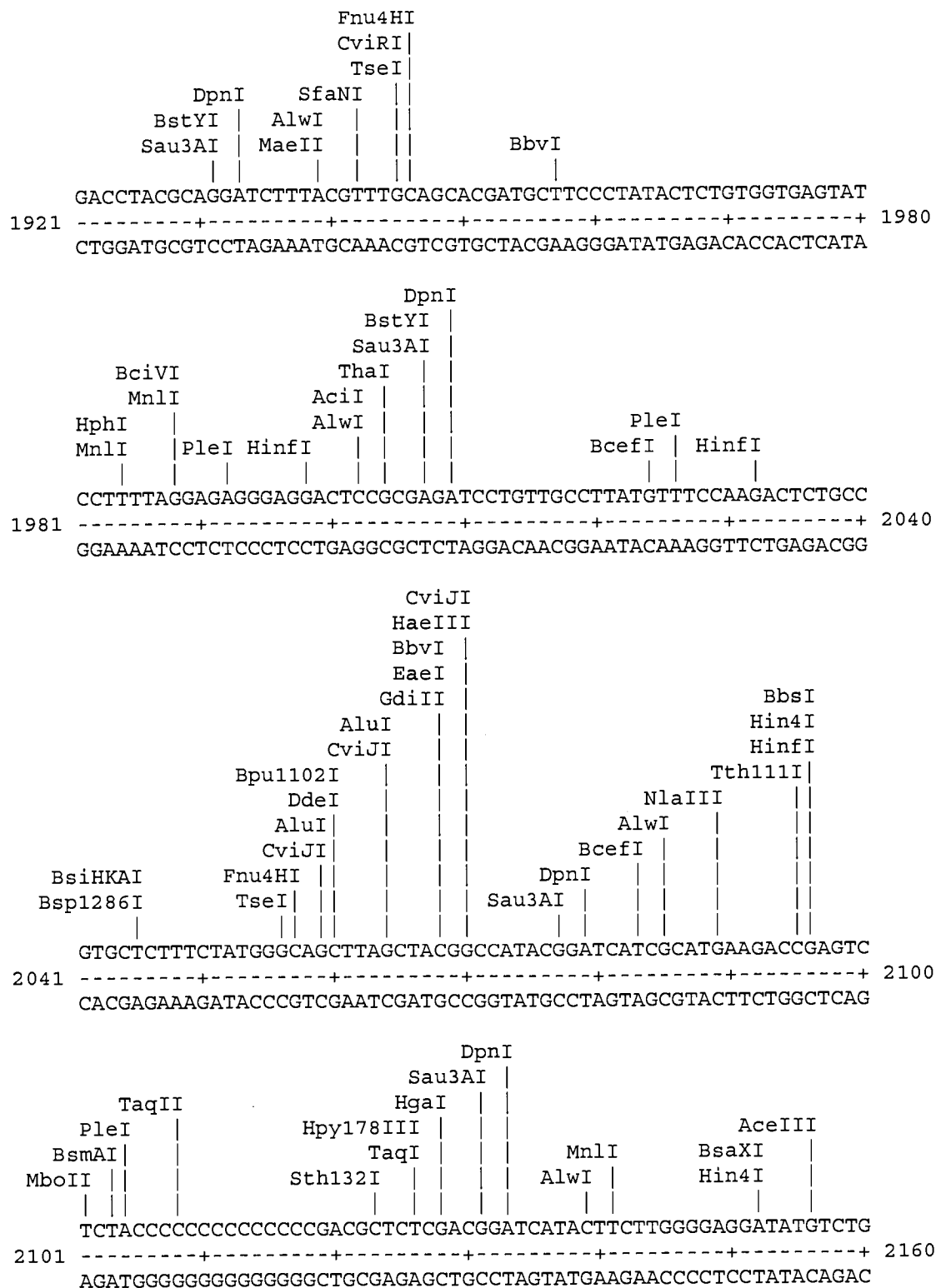


Fig. 2 (con't)



Fnu4HI  
 TauI  
 AcII  
 MspAII  
 MnlII  
 CviJI  
 AluI  
 SmlI  
 XhoI  
 BpmI  
 TaqI  
 AvaI  
 GGCTGGAGAGCTGGGAACCTCGAGTTGCTGTTGAAAATACCAGCGGCAGAGGATTTTCCA  
 2161 -----+-----+-----+-----+-----+-----+-----+ 2220  
 CCGACCTCTCGACCCTTGAGCTCAACGACAACCTTTTATGGTCGCCGTCTCCTAAAAAGGT  
 AluI  
 CviJI  
 MwoI  
 Cac8I  
 MwoI  
 CviJI  
 RsaI  
 TatI  
 AGAGTACACTCCATTTGTAAAAGTCCAAGCTGTTTACGCTCGCCAAGATAGCTTTGTAGA  
 2221 -----+-----+-----+-----+-----+-----+-----+ 2280  
 TCTCATGTGAGGTAAACATTTTCAGGTTTCACAAATGCGAGCGGTTCTATCGAAACATCT  
 BsaBI  
 HinfI  
 TfiI  
 CjePI  
 SfaNI  
 HinfI  
 TfiI  
 AluI  
 BfaI  
 CviJI  
 Hpy178III  
 ACTAGGAGCTATCAGTCGTGATTTTAGTGATTTCGCATCTTTATAACCTTGCGATTCCCTCT  
 2281 -----+-----+-----+-----+-----+-----+-----+ 2340  
 TGATCCTCGATAGTCAGCACTAAAATCACTAAGCGTAGAAATATTGGAACGCTAAGGAGA  
 MnlI  
 HinfI  
 TfiI  
 CjePI  
 TaaI  
 CviRI  
 SspI  
 NlaIII  
 BpmI  
 Tth111III  
 Hin4I  
 TGGAATCAAGTTAGAGAAAACGGTTTGCAGAGCAATATTATCATGTTGTAGCGATGTATTC  
 2341 -----+-----+-----+-----+-----+-----+-----+ 2400  
 ACCTTAGTTCAATCTCTTTGCCAAACGTCTCGTTATAATAGTACAACATCGCTACATAAG  
 BslI  
 BsaJI  
 StyI  
 MaeIII  
 Pfl1108I  
 RsaI  
 MmeI  
 Hpy178III  
 TCCAGATGTTTGTCTAGTAACCCCCAAATGTACGACTACCCTACTTTCCAACCAAGGGAG  
 2401 -----+-----+-----+-----+-----+-----+-----+ 2460  
 AGGTCTACAAACAGCATCATTGGGGTTTACATGCTGATGGGATGAAAGGTTGGTTCCTC  
 Bsu36I  
 DdeI  
 CviJI  
 HaeI  
 HaeIII  
 StuI  
 AceIII  
 NspV  
 TaqI  
 MboII  
 MmeI  
 BbsI  
 DdeI  
 CviJI  
 MwoI  
 TTGGAAGACCAAAGGTTCGAACTTAGCAAGACAGGCTGGTATTGTTTCAGGCCTCAGGTTT  
 2461 -----+-----+-----+-----+-----+-----+-----+ 2520  
 AACCTTCTGGTTTCCAAGCTTGAATCGTTCTGTCCGACCATAACAAGTCCGGAGTCCAAA

Fig. 2 (con't)

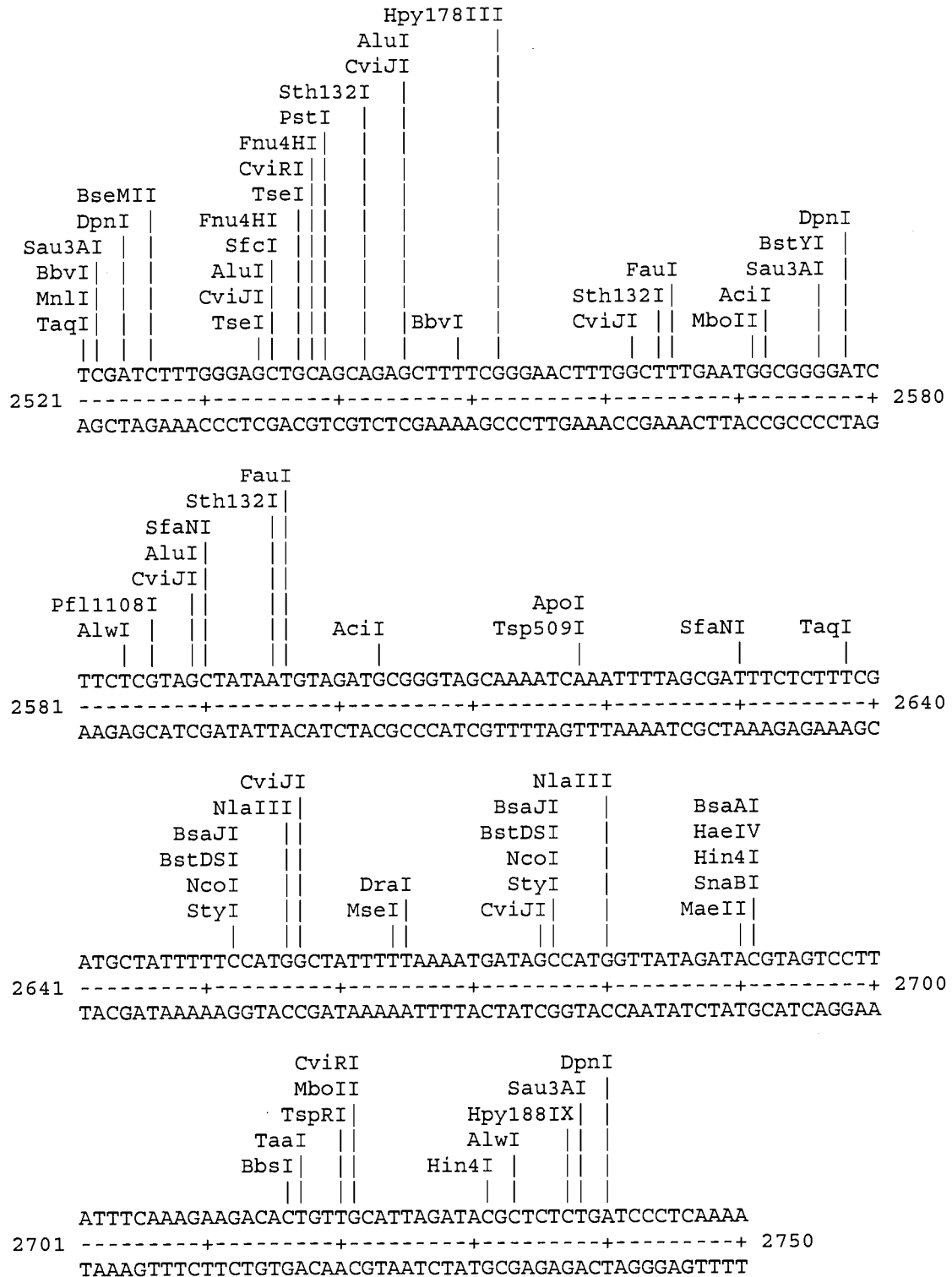


Figure 3: CPN100421

```

ctcctgtccc tcgcttgctc aacctacccc tcctccctcg aattctaadc ctttgaacgt 60
agtagaacag cctgttgctg catcgtcagt gccttcctac atg ccc cca ctg aat 115
                                     Met Pro Pro Leu Asn
                                     1 5
gct gat gat gtt ctc cct aga gac cat ctg tca gat gga agt ttc tca 163
Ala Asp Asp Val Leu Pro Arg Asp His Leu Ser Asp Gly Ser Phe Ser
                                     10 15 20
gat acg tat cca gac att aca acg caa gcg atc atc tta att ttc ttg 211
Asp Thr Tyr Pro Asp Ile Thr Thr Gln Ala Ile Ile Leu Ile Phe Leu
                                     25 30 35
gcc cta tcg cct ttc ctg gtc atg ttg ctc act tcg tat cta aag att 259
Ala Leu Ser Pro Phe Leu Val Met Leu Leu Thr Ser Tyr Leu Lys Ile
                                     40 45 50
atc att act tta gtc tta tta cgt aac gcc tta gga gta caa caa aca 307
Ile Ile Thr Leu Val Leu Leu Arg Asn Ala Leu Gly Val Gln Gln Thr
                                     55 60 65
cct ccc agt caa gtc ctc aat ggg att gca ctc atc cta tct att tat 355
Pro Pro Ser Gln Val Leu Asn Gly Ile Ala Leu Ile Leu Ser Ile Tyr
                                     70 75 80 85
gtg atg ttc ccc acg gga gtg gct atg tat aaa gat gct cgc aag gaa 403
Val Met Phe Pro Thr Gly Val Ala Met Tyr Lys Asp Ala Arg Lys Glu
                                     90 95 100
atc gaa gcc aat acc att cct caa agc ctc ttc act gca gaa ggt gct 451
Ile Glu Ala Asn Thr Ile Pro Gln Ser Leu Phe Thr Ala Glu Gly Ala
                                     105 110 115
gaa aca gtg ttt gtc gct tta aac aaa tct aaa gaa cct ttg cgc tct 499
Glu Thr Val Phe Val Ala Leu Asn Lys Ser Lys Glu Pro Leu Arg Ser
                                     120 125 130
ttc tta att cgc aac act cca aaa gca caa att caa agc ttt tac aag 547
Phe Leu Ile Arg Asn Thr Pro Lys Ala Gln Ile Gln Ser Phe Tyr Lys
                                     135 140 145
atc tca cag aaa acc ttc cct tcg gaa att cga gcg cac ctc act gcc 595
Ile Ser Gln Lys Thr Phe Pro Ser Glu Ile Arg Ala His Leu Thr Ala
                                     150 155 160 165
tcc gac ttt gta atc att att cct gct ttt att atg ggt cag ata aaa 643
Ser Asp Phe Val Ile Ile Ile Pro Ala Phe Ile Met Gly Gln Ile Lys
                                     170 175 180
aat gct ttc gaa att gga gtc ttg atc tat cta cct ttc ttt gtt att 691
Asn Ala Phe Glu Ile Gly Val Leu Ile Tyr Leu Pro Phe Phe Val Ile
                                     185 190 195

```

Fig. 3 (con't)

gat tta gtg act gct aac gtt ctt gta gcg atg cag atg atg atg tta	739
Asp Leu Val Thr Ala Asn Val Leu Val Ala Met Gln Met Met Met Leu	
200 205 210	
tcc cct cta tcg att tcg tta cct tta aag tta ctt ttg atc gtc atg	787
Ser Pro Leu Ser Ile Ser Leu Pro Leu Lys Leu Leu Leu Ile Val Met	
215 220 225	
gta gac gga tgg aca tta ctg ctc caa ggg ctt atg atc agc ttt aaa	835
Val Asp Gly Trp Thr Leu Leu Leu Gln Gly Leu Met Ile Ser Phe Lys	
230 235 240 245	
taaggacacg tgccgtgtta gcattttttcg caactagttt caaatctggt ctttttgagt	895
actcctacca atcattatta cttattttga ttgtttcggc acctcccatc atcttagctt	955
ccatagtcgg gattatggtt gcgatcttcc aagccgcaac acaaa	1000

Figure 4 (RY-34)

## Restriction enzyme analysis of CP100421

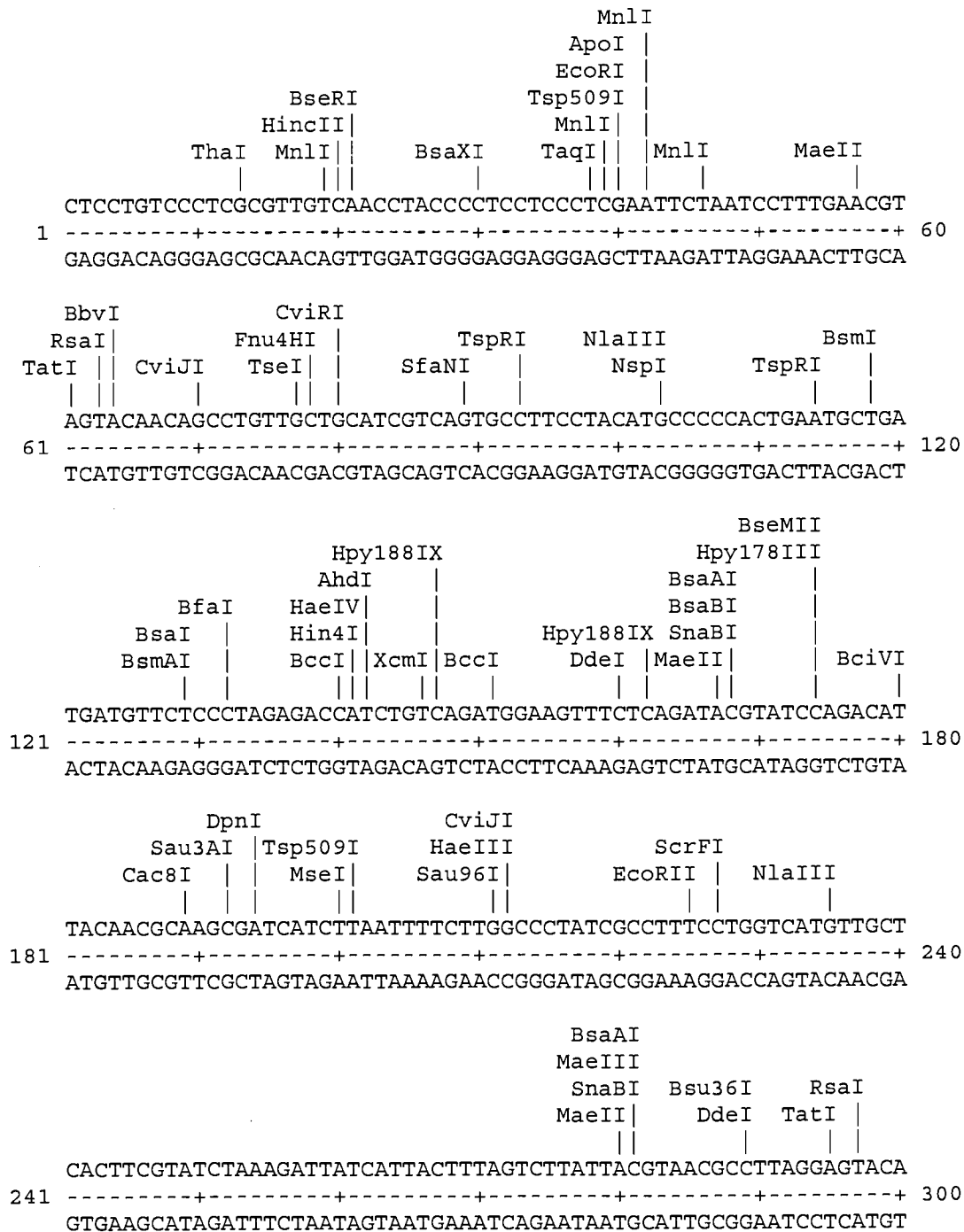


Fig. 4 (con't)

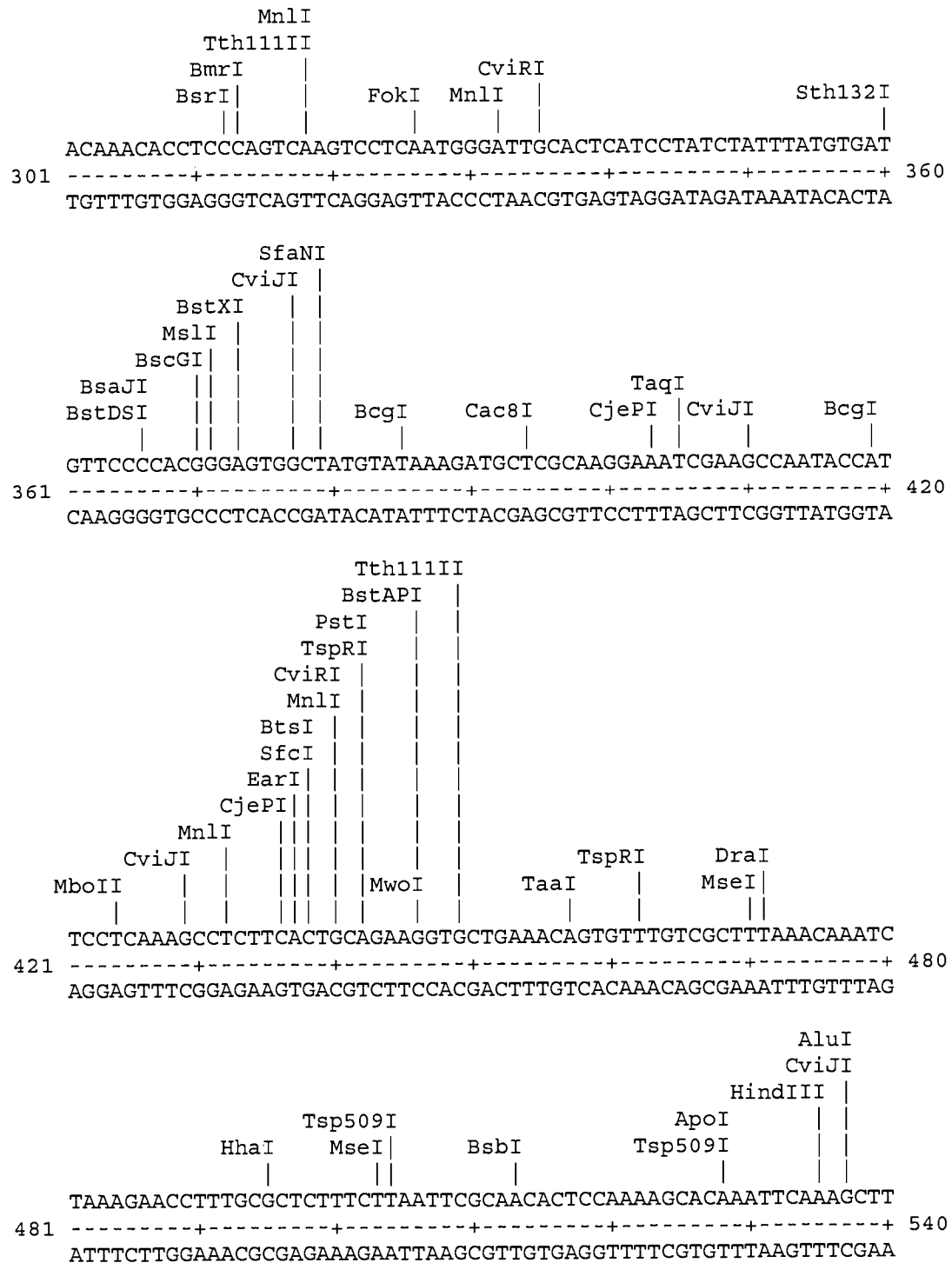




Fig. 4 (con't)

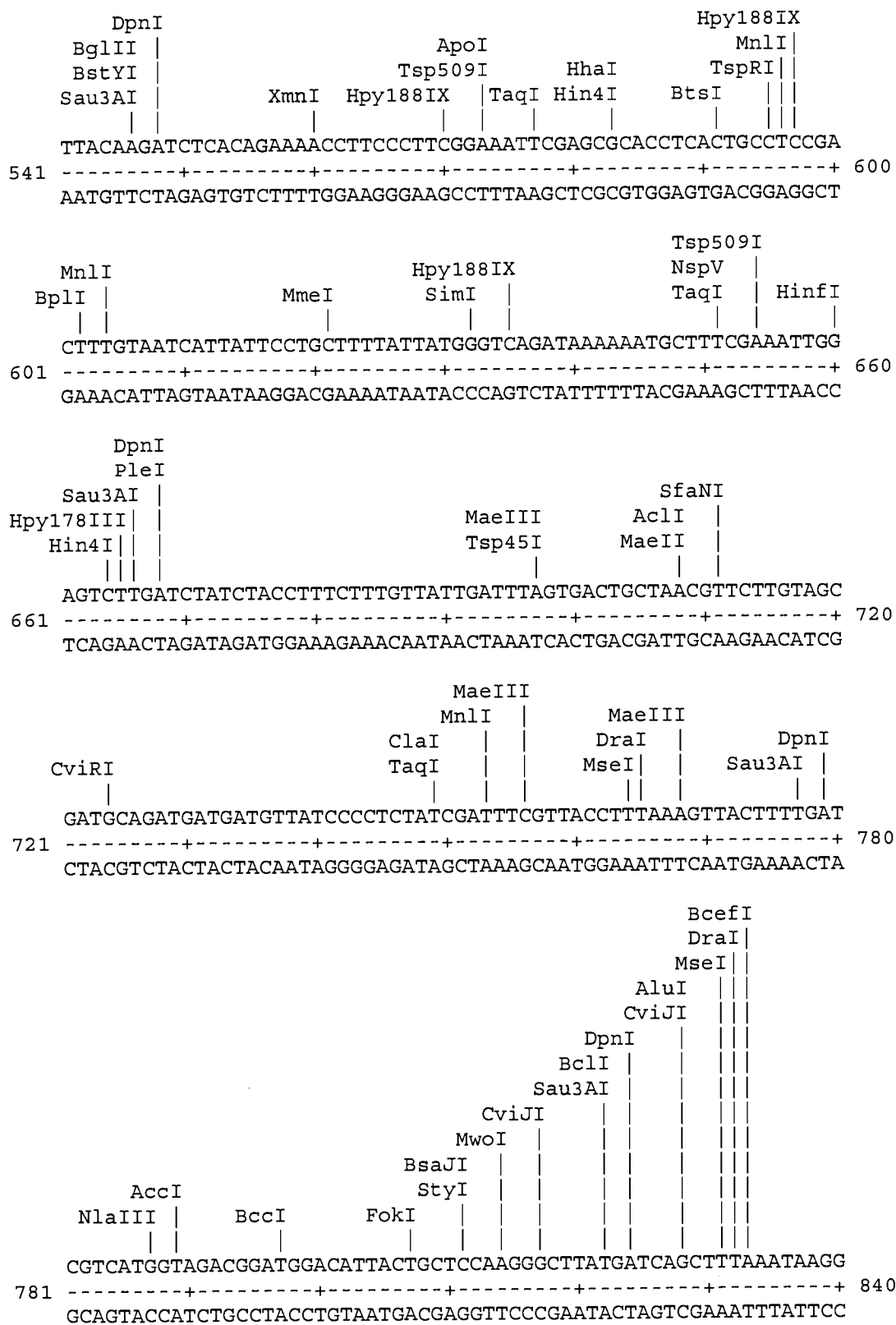


Fig. 4 (con't)



Figure 5:

```

tagctttata caaagtatag aaaaataaca cgacaataaa aggagcgggtg ttttctcttc 60

tgaggtaa at cagcctcaaa gatactacgc catagtaaag atg aag ttt ttt agc 115
              Met Lys Phe Phe Ser
              1 5

tta att ttt aaa gat gat gat gtc tcc cca aat aag aag gtt tta tct 163
Leu Ile Phe Lys Asp Asp Asp Val Ser Pro Asn Lys Lys Val Leu Ser
              10 15 20

cct gaa gct ttc tct gct ttc ctt gat gcc aaa gag ctg tta gaa aaa 211
Pro Glu Ala Phe Ser Ala Phe Leu Asp Ala Lys Glu Leu Leu Glu Lys
              25 30 35

aca aaa gcc gat agc gaa gcc tat gtt gca gag aca gaa caa aag tgt 259
Thr Lys Ala Asp Ser Glu Ala Tyr Val Ala Glu Thr Glu Gln Lys Cys
              40 45 50

gca caa att cgt caa gaa gct aaa gat caa gga ttt aaa gag gga tct 307
Ala Gln Ile Arg Gln Glu Ala Lys Asp Gln Gly Phe Lys Glu Gly Ser
              55 60 65

gaa tcc tgg agc aag caa att gct ttc tta gaa gaa gaa act aaa aat 355
Glu Ser Trp Ser Lys Gln Ile Ala Phe Leu Glu Glu Glu Thr Lys Asn
              70 75 80 85

cta cgc ata aga gta cgc gag gcc ttg gtt cct ctg gca att gcg agt 403
Leu Arg Ile Arg Val Arg Glu Ala Leu Val Pro Leu Ala Ile Ala Ser
              90 95 100

gtg agg aaa atc att ggg aag gaa ctc gaa tta cat cct gaa act att 451
Val Arg Lys Ile Ile Gly Lys Glu Leu Glu Leu His Pro Glu Thr Ile
              105 110 115

gtc tct att att tct caa gca ttg aaa gag ctc aca caa aat aaa cat 499
Val Ser Ile Ile Ser Gln Ala Leu Lys Glu Leu Thr Gln Asn Lys His
              120 125 130

atc att atc tct gtc aat ccc aaa gat tta cct ctt gtt gag aaa agt 547
Ile Ile Ile Ser Val Asn Pro Lys Asp Leu Pro Leu Val Glu Lys Ser
              135 140 145

cgt cct gaa ctc aag aac atc gtg gag tat gct gac tcc tta att ctt 595
Arg Pro Glu Leu Lys Asn Ile Val Glu Tyr Ala Asp Ser Leu Ile Leu
              150 155 160 165

aca gca aaa cct gat gtt act cct ggg ggt tgc att atc gag act gaa 643
Thr Ala Lys Pro Asp Val Thr Pro Gly Gly Cys Ile Ile Glu Thr Glu
              170 175 180

gca ggg atc atc aat gcg cag ctt gat gta caa tta gat gcc tta gaa 691
Ala Gly Ile Ile Asn Ala Gln Leu Asp Val Gln Leu Asp Ala Leu Glu
              185 190 195

```

Fig. 5 (con't)

```

aaa gct ttc tcg act ata cta aaa gcg aag aac cct gta gac gag cca 739
Lys Ala Phe Ser Thr Ile Leu Lys Ala Lys Asn Pro Val Asp Glu Pro
      200                205                210

tct gag act tca tca tcc acg gat tct tct tct tta tct aat gat cag 787
Ser Glu Thr Ser Ser Ser Thr Asp Ser Ser Ser Leu Ser Asn Asp Gln
      215                220                225

gat aag aaa gaa taaaggtatt cactattatg cgatccattt ttcgattttc 839
Asp Lys Lys Glu
230

cctttgtttt tttagcgtga gcgtctcatg ctgatttgct gacgccagtc tatatgaaaa 899

c 900

```

Figure 6 (RY-35)

Restriction analysis of CPN100422

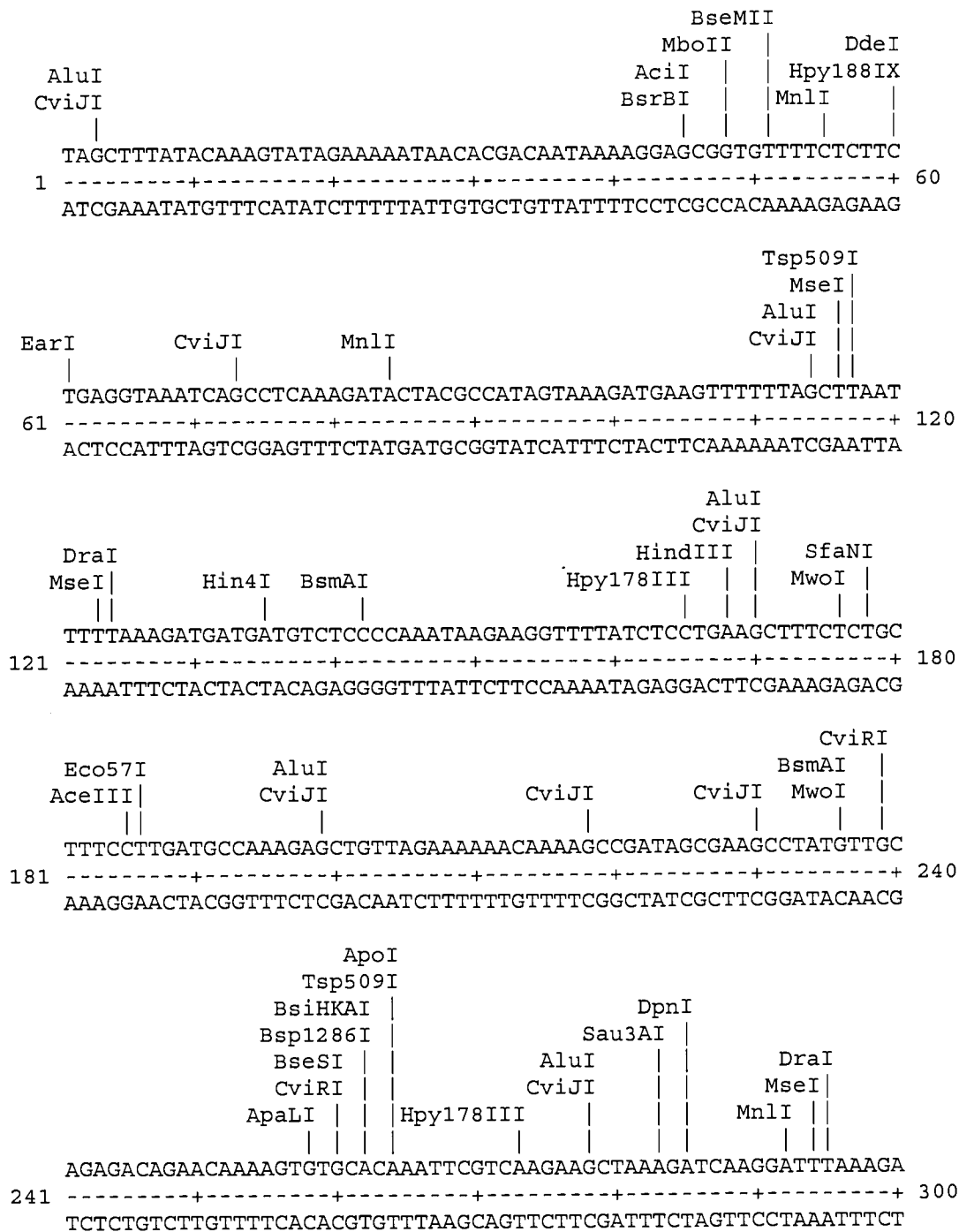


Fig. 6 (con't)

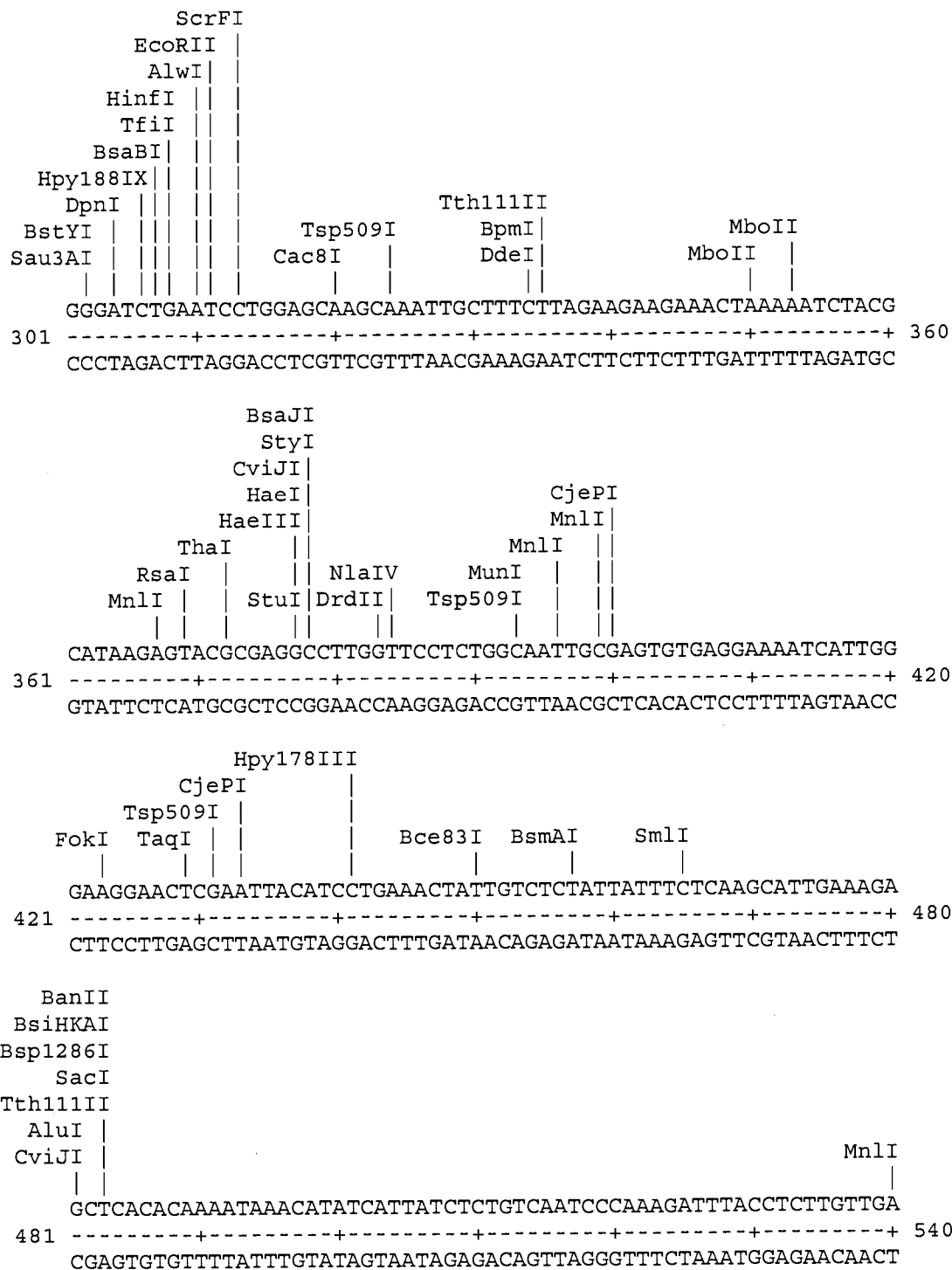
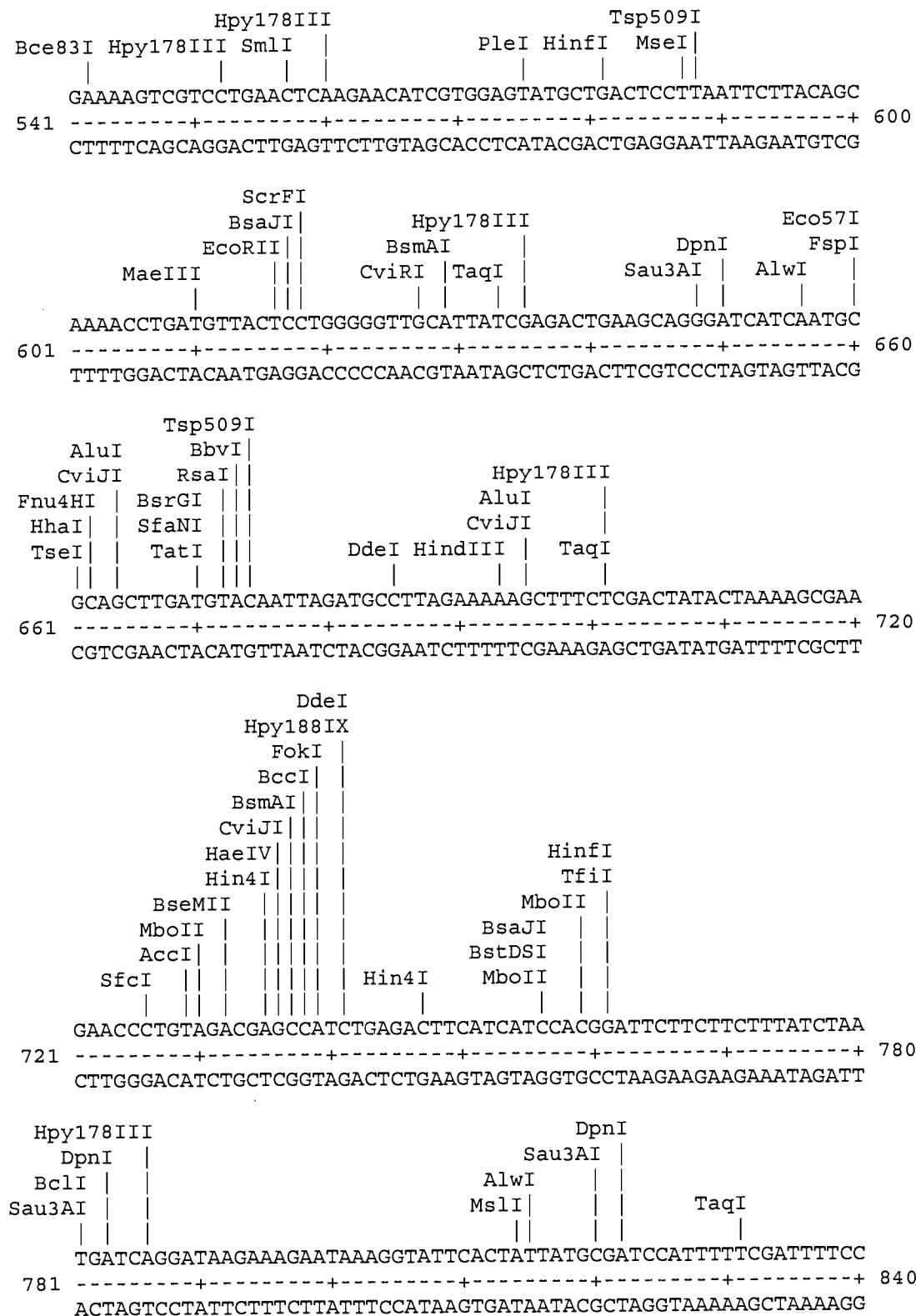


Fig. 6 (con't)



```

                                NlaIII
                                |
          Bpu1102I      BsmAI  |      BsrI
          |             |      |
      HgaI             |      |      PshAI
      |             |      |      |
BseMII |      DdeI  |      |      BsaHI |      HgaI
      | |      |      |      |      |
CTTTGTTTTTTTACGCTGAGCGTCTCATGCTGATTTGCTGACGCCAGTCTATATGAAAAC
841 -----+-----+-----+-----+-----+-----+-----+ 900
GAAACAAAAAAATGCGACTCGCAGAGTACGACTAAACGACTGCGGTGAGATATACTTTTG

```



Figure 7: CPN 100424

```

tggtcgcgat tggcactaat cccccctttt gttatggtga ataaaaaggt atgcgtggat 60

tatggttcgt cgatctatctt ctttttgctt gttcttttcta atg aca ttg ctg tgc 115
                                         Met Thr Leu Leu Cys
                                         1                               5

tgt aca agc tgt aac agc agg tct cta att gtg cac ggt ctt cct ggc 163
Cys Thr Ser Cys Asn Ser Arg Ser Leu Ile Val His Gly Leu Pro Gly
                        10                               15                               20

aga gaa gcg aat gag att gtg gtg ctt ttg gta agc aaa ggg gtg gct 211
Arg Glu Ala Asn Glu Ile Val Val Leu Leu Val Ser Lys Gly Val Ala
                        25                               30                               35

gca caa aaa ttg cct caa gct gca gcg gct aca gcc gga gca gct act 259
Ala Gln Lys Leu Pro Gln Ala Ala Ala Ala Thr Ala Gly Ala Ala Thr
                        40                               45                               50

gag caa atg tgg gat atc gcg gtt ccg tca gca caa atc aca gag gcc 307
Glu Gln Met Trp Asp Ile Ala Val Pro Ser Ala Gln Ile Thr Glu Ala
                        55                               60                               65

ctt gcc att cta aat caa gcg ggt ctt cca cgt atg aaa ggg aca agc 355
Leu Ala Ile Leu Asn Gln Ala Gly Leu Pro Arg Met Lys Gly Thr Ser
                        70                               75                               80                               85

ctg tta gat ctt ttt gca aaa caa ggt ctt gtt cct tcc gag ctt cag 403
Leu Leu Asp Leu Phe Ala Lys Gln Gly Leu Val Pro Ser Glu Leu Gln
                        90                               95                               100

gaa aaa atc cgt tat caa gaa ggc tta tca gaa cag atg gcc tct acg 451
Glu Lys Ile Arg Tyr Gln Glu Gly Leu Ser Glu Gln Met Ala Ser Thr
                        105                               110                               115

att aga aaa atg gat ggc gtt gtc gat gcc tca gta cag att tcc ttc 499
Ile Arg Lys Met Asp Gly Val Val Asp Ala Ser Val Gln Ile Ser Phe
                        120                               125                               130

act aca gaa aat gaa gat aat ctt cct tta aca gcc tct gtg tat att 547
Thr Thr Glu Asn Glu Asp Asn Leu Pro Leu Thr Ala Ser Val Tyr Ile
                        135                               140                               145

aag cat cga ggg gtt ttg gac aat ccg aac agc att atg gtt tcc aaa 595
Lys His Arg Gly Val Leu Asp Asn Pro Asn Ser Ile Met Val Ser Lys
                        150                               155                               160                               165

att aag cgc ctt att gca agt gct gtt cca gga ctt gtg cca gag aac 643
Ile Lys Arg Leu Ile Ala Ser Ala Val Pro Gly Leu Val Pro Glu Asn
                        170                               175                               180

gtc tct gta gtg agc gat cgc gca gct tat agt gat att aca att aat 691
Val Ser Val Val Ser Asp Arg Ala Ala Tyr Ser Asp Ile Thr Ile Asn
                        185                               190                               195

```

Fig. 7 (con't)

ggt cct tgg gga tta aca gaa gaa atc gat tat gtt tct gtt tgg ggt	739
Gly Pro Trp Gly Leu Thr Glu Glu Ile Asp Tyr Val Ser Val Trp Gly	
200 205 210	
att att ctt gcg aag tct tcg ctc acc aaa ttc cgt ctc att ttt tat	787
Ile Ile Leu Ala Lys Ser Ser Leu Thr Lys Phe Arg Leu Ile Phe Tyr	
215 220 225	
gtc ttg att ctc att tta ttt gtt att tct tgt ggt ctc ctt tgg gtc	835
Val Leu Ile Leu Ile Leu Phe Val Ile Ser Cys Gly Leu Leu Trp Val	
230 235 240 245	
att tgg aaa act cat act ctc att atg act atg gga ggt aca aaa ggg	883
Ile Trp Lys Thr His Thr Leu Ile Met Thr Met Gly Gly Thr Lys Gly	
250 255 260	
ttc ttc aac cct aca cca tat aca aag aat gcc ttg gaa gcc aag aaa	931
Phe Phe Asn Pro Thr Pro Tyr Thr Lys Asn Ala Leu Glu Ala Lys Lys	
265 270 275	
gcc gag gga gca gct gct gac aaa gag aaa aaa gaa gat gca gat tca	979
Ala Glu Gly Ala Ala Ala Asp Lys Glu Lys Lys Glu Asp Ala Asp Ser	
280 285 290	
cag ggg gaa agc aaa aat gcg gaa acc agt gat aaa gac tct agt gat	1027
Gln Gly Glu Ser Lys Asn Ala Glu Thr Ser Asp Lys Asp Ser Ser Asp	
295 300 305	
aaa gat gct cca gaa gga agc aat gaa att gag ggt gct tagtgactgc	1076
Lys Asp Ala Pro Glu Gly Ser Asn Glu Ile Glu Gly Ala	
310 315 320	
caacactttt ggaactctag acatcttgat gaagcactcc aaggaagatg acctctccag	1136
gtttcttcct aaaaatcttc ttgttgaatc tctcatccc gaagaaatcc ctttaaaatc	1196
ttta	1200

Figure 8 (RY-36)

## Restriction analysis of CPN100424



Fig. 8 (con't)

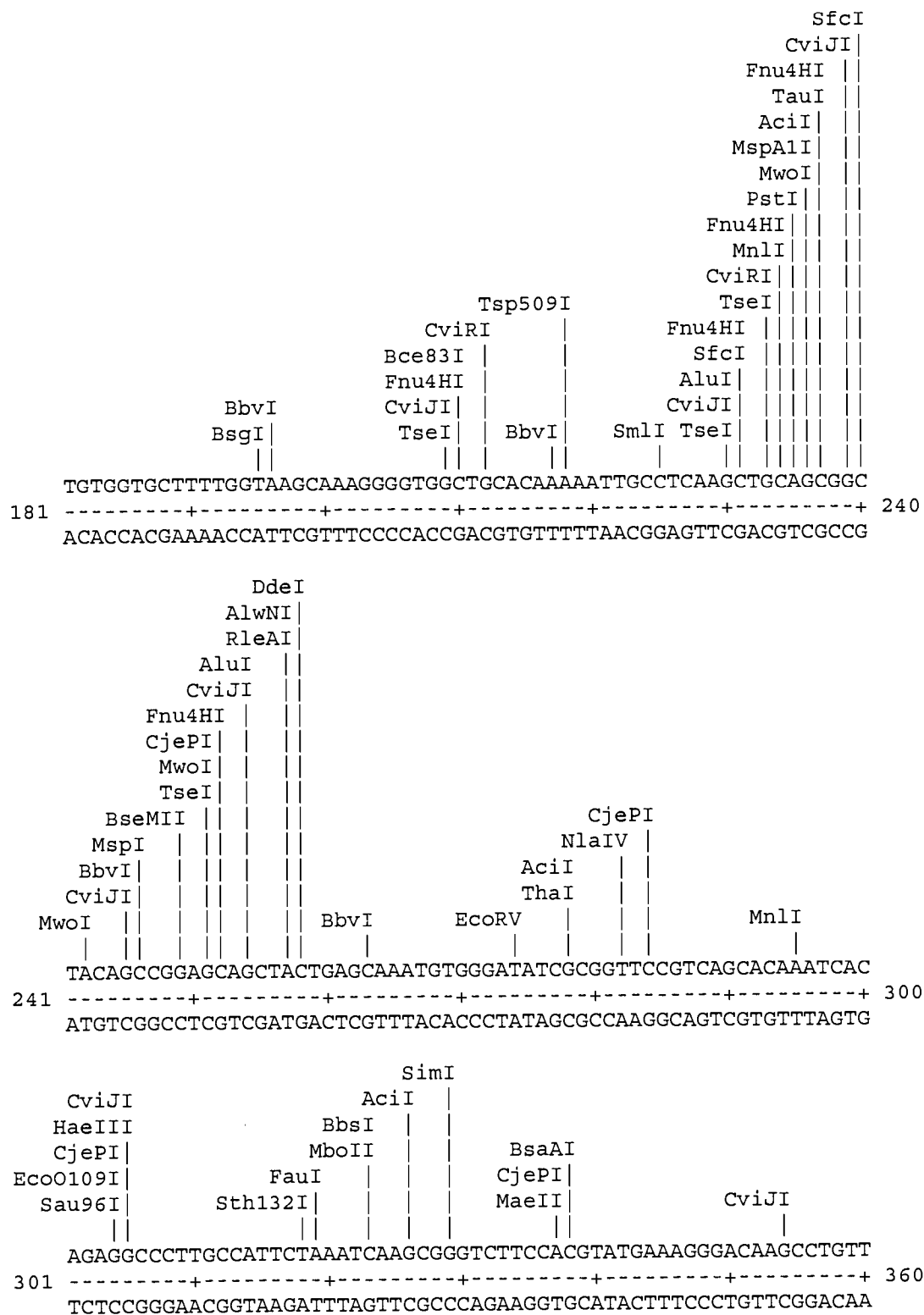


Fig. 8 (con't)

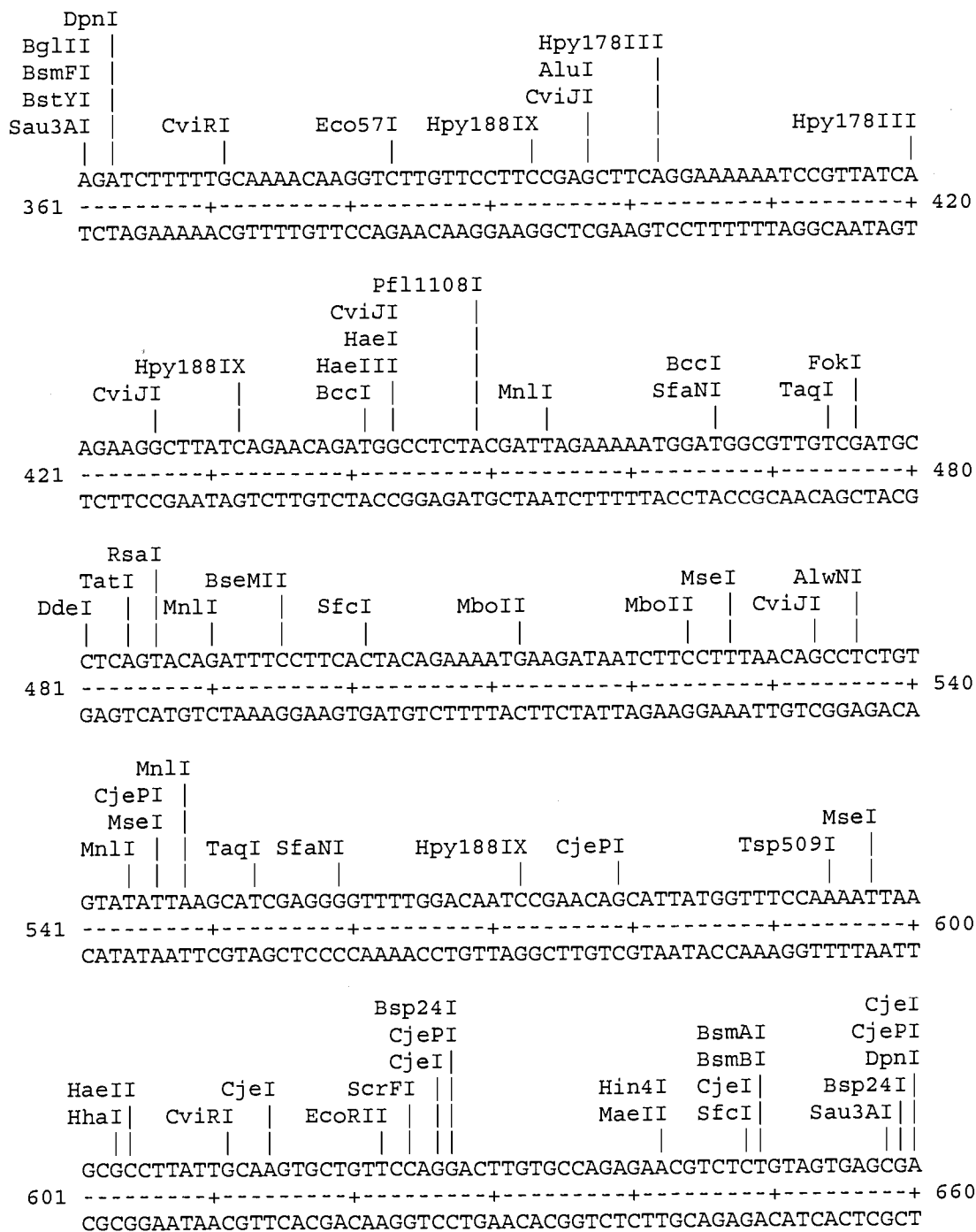


Fig. 8 (con't)

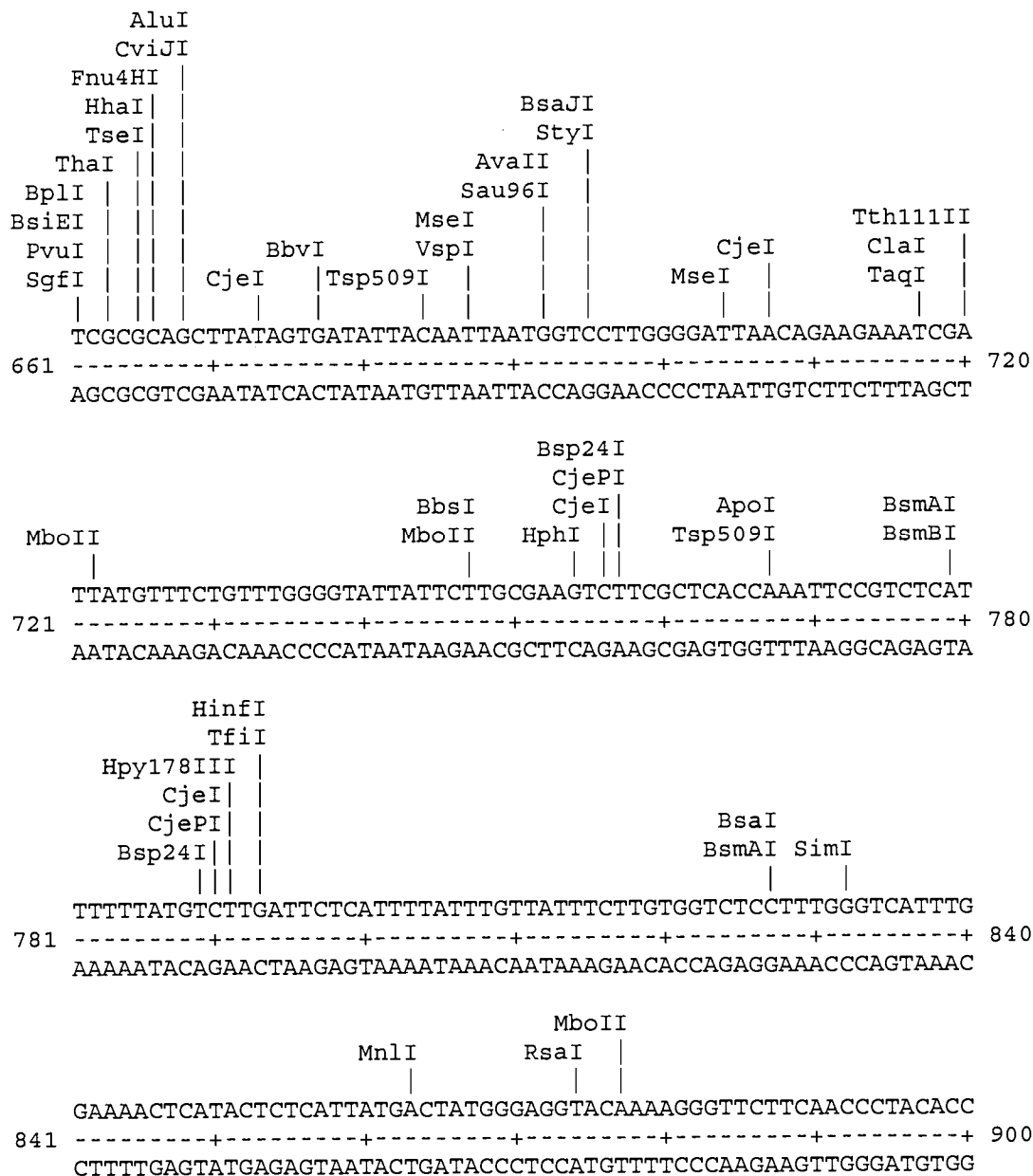


Fig. 8 (con't)

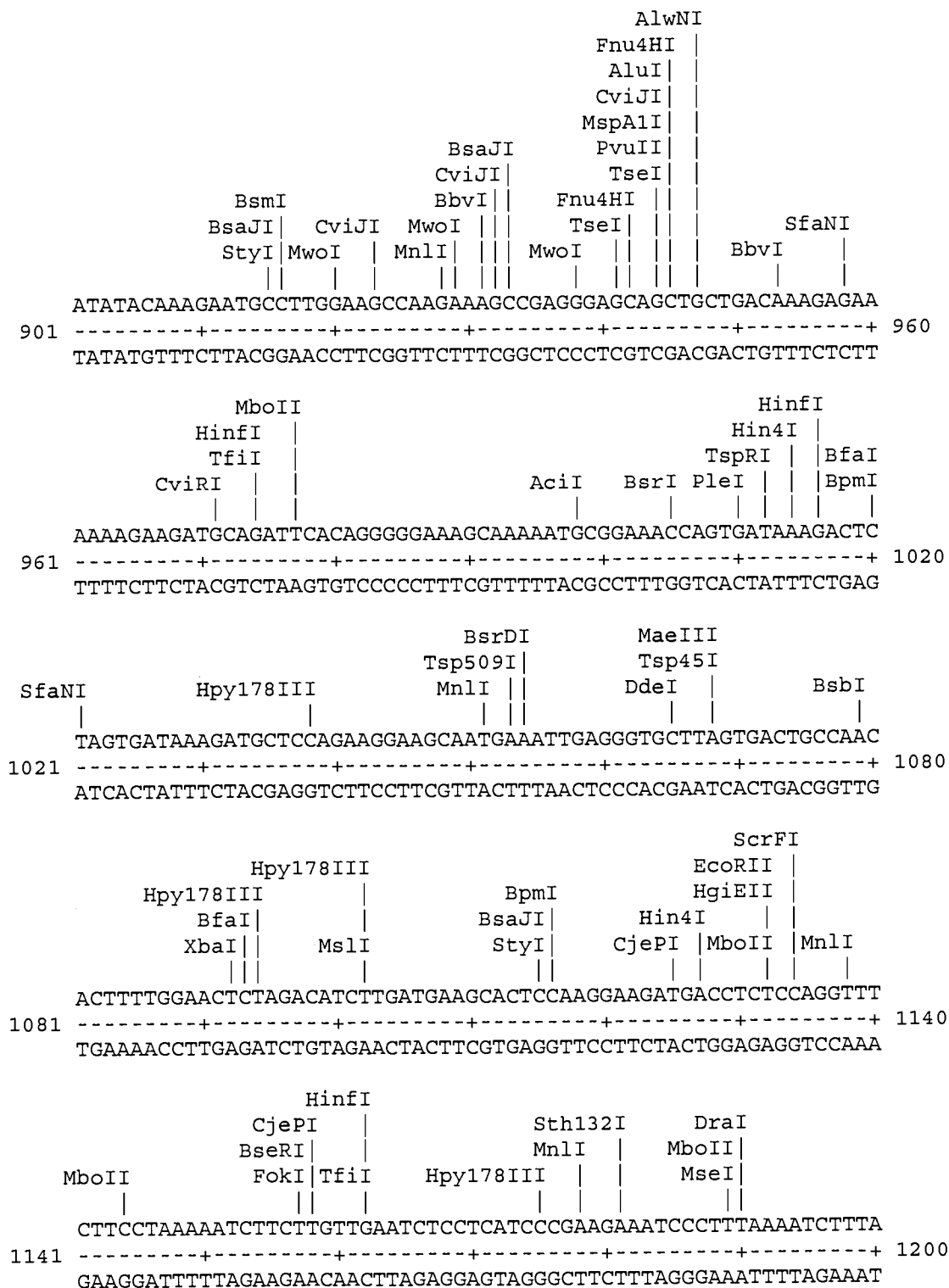


Figure 9: CPN100426

```

ttgaacccta tggaaatgta tcttatttgt gctgggctat atttcttaat gacaacatca 60
ttttcctgta tttctagggt atcagaaaaag agaaggagtt atg aca att aga gtc 115
                                         Met Thr Ile Arg Val
                                         1           5

cga aac ctt gcc tac tct gta aat aag aaa aag att cta gat ggt gta 163
Arg Asn Leu Ala Tyr Ser Val Asn Lys Lys Lys Ile Leu Asp Gly Val
              10              15              20

act ttt tct tta gag cga ggg cac att aca ctg ttt gtt ggg aag agt 211
Thr Phe Ser Leu Glu Arg Gly His Ile Thr Leu Phe Val Gly Lys Ser
              25              30              35

ggg tca gga aaa aca atg att tta cgt gct ttg gcg ggc tta gtc cag 259
Gly Ser Gly Lys Thr Met Ile Leu Arg Ala Leu Ala Gly Leu Val Gln
              40              45              50

ccc act caa gga gat att tgg att gaa ggg gag gct cca gct cta gtt 307
Pro Thr Gln Gly Asp Ile Trp Ile Glu Gly Glu Ala Pro Ala Leu Val
              55              60              65

ttc caa caa ccc gag tta ttt tcc cat atg aca gta tta gga aat tgc 355
Phe Gln Gln Pro Glu Leu Phe Ser His Met Thr Val Leu Gly Asn Cys
              70              75              80              85

acc cat cca caa atc cat atc aag ggt cgt agt acc gaa gaa gct cga 403
Thr His Pro Gln Ile His Ile Lys Gly Arg Ser Thr Glu Glu Ala Arg
              90              95              100

gaa aag gcg ttc gag ctt tta cat ttg ttg gat att gaa gag gtt gct 451
Glu Lys Ala Phe Glu Leu Leu His Leu Leu Asp Ile Glu Glu Val Ala
              105              110              115

aag aat tat cct gac cag ctc tct ggg gga caa aaa caa cgt gtg gct 499
Lys Asn Tyr Pro Asp Gln Leu Ser Gly Gly Gln Lys Gln Arg Val Ala
              120              125              130

att gta cgt tct tta tgt atg gat aaa cat aca tta ctt ttt gat gaa 547
Ile Val Arg Ser Leu Cys Met Asp Lys His Thr Leu Leu Phe Asp Glu
              135              140              145

cct aca tcg gct tta gat cct ttt gct acg gca tcg ttc cga cat ctt 595
Pro Thr Ser Ala Leu Asp Pro Phe Ala Thr Ala Ser Phe Arg His Leu
              150              155              160              165

tta gaa aca ctt cga gac cag gaa ctg act gta ggg tta act act cat 643
Leu Glu Thr Leu Arg Asp Gln Glu Leu Thr Val Gly Leu Thr Thr His
              170              175              180

gac atg caa ttt gtt cat agt tgt ttg gat cgt atc tat ctt ata gat 691
Asp Met Gln Phe Val His Ser Cys Leu Asp Arg Ile Tyr Leu Ile Asp
              185              190              195

```



Fig. 9 (con't)

```
caa gga act gtt gcg ggg gtc tat gac aag cgt gac gga gag ctc gat 739
Gln Gly Thr Val Ala Gly Val Tyr Asp Lys Arg Asp Gly Glu Leu Asp
      200                      205                      210

tct ggt cat cca tta tcg aaa tat atc cac tct gct caa taggactaca 788
Ser Gly His Pro Leu Ser Lys Tyr Ile His Ser Ala Gln
      215                      220                      225

gctgctagag cagctgtagt gatactttag aatcctgacc agtggcagga atgagcggca 848

tg 850
```

Figure 10 (RY-37)  
Restriction enzyme analysis of CPN100426

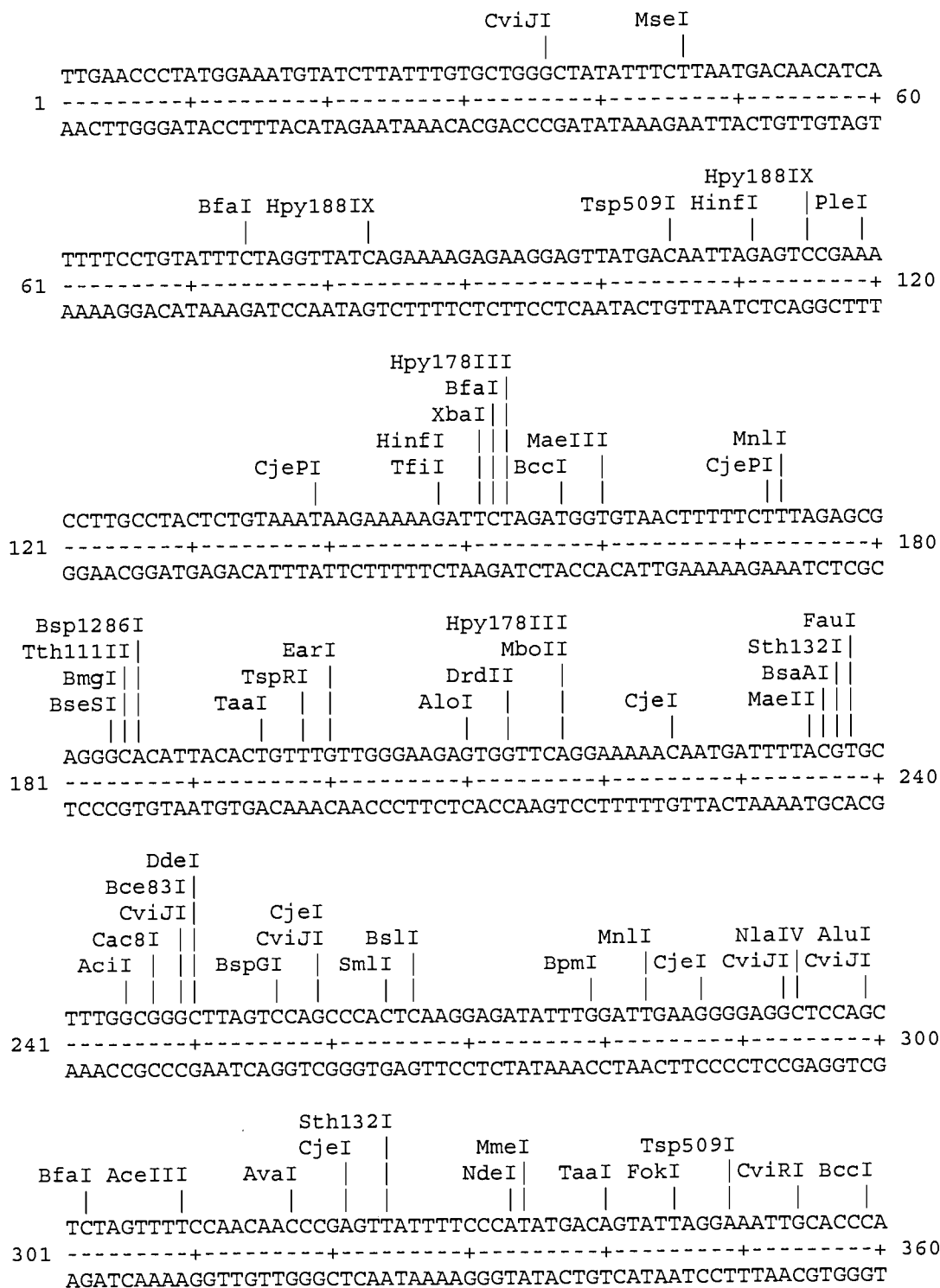


Fig. 10 (con't)

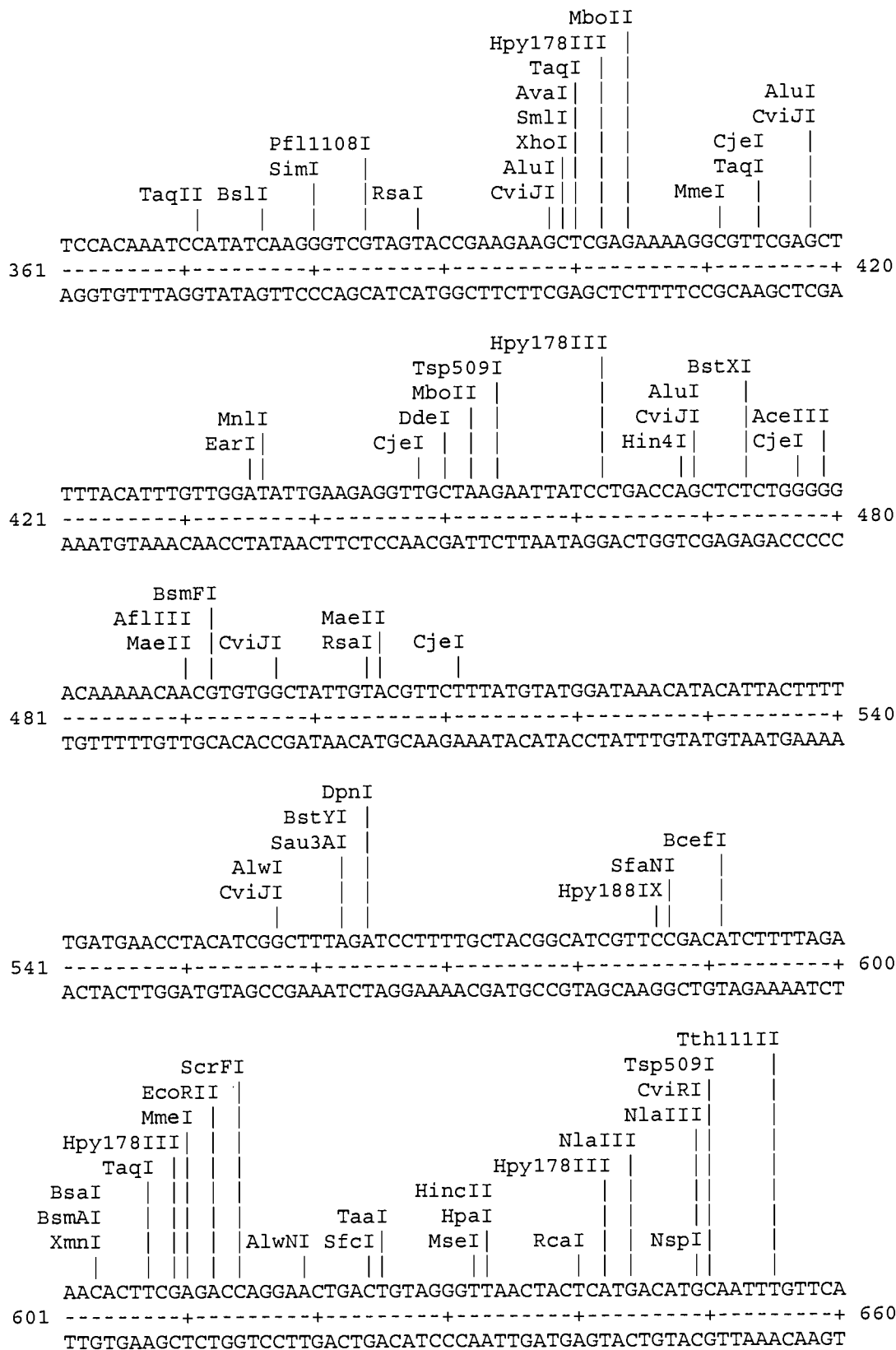


Fig. 10 (con't)

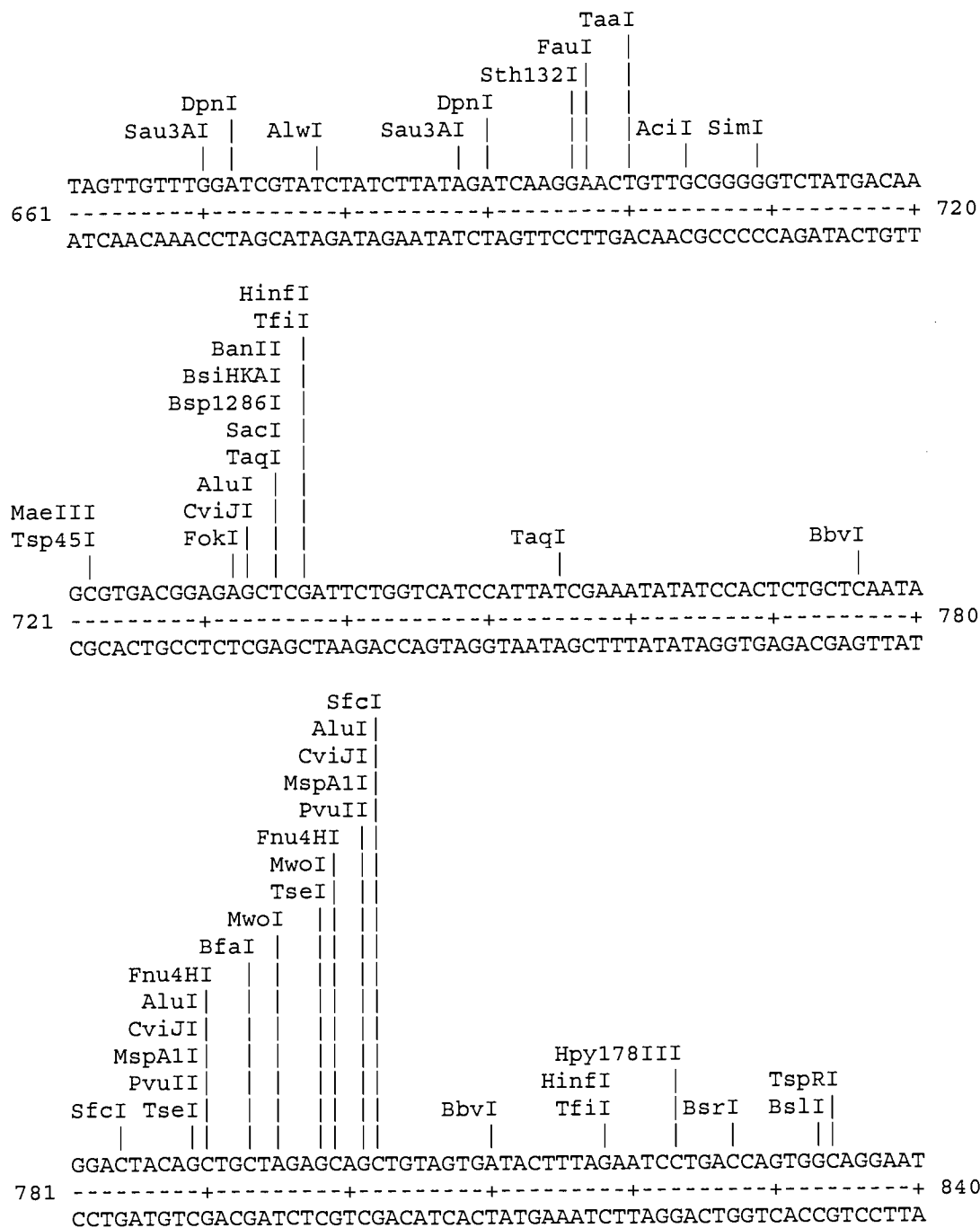


Fig. 10 (con't)

```

      Fnu4HI
      TauI
      AclI |
      BsrBI | NlaIII
          ||      |
          GAGCGGCATG
841 -----+ 850
      CTCGCCGTAC

```

Figure 11: CPN100508

```

ctctgattta tggtaattct ttatttttcag agccgtcaag tcctttctat tctgttgaat 60
ttcctaataa cgtaagtaat aaacaatcaa aagtcgcgat atg aaa aga cct ttt 115
                                     Met Lys Arg Pro Phe
                                     1           5

ttt acc tat cta tgc atc atc ttc tac gga tct tgt gca tcg tta tct 163
Phe Thr Tyr Leu Cys Ile Ile Phe Tyr Gly Ser Cys Ala Ser Leu Ser
                        10                15                20

tta cat gca gga ctc tct ttc cca gaa gta cgt gga gct acg gct gct 211
Leu His Ala Gly Leu Ser Phe Pro Glu Val Arg Gly Ala Thr Ala Ala
                        25                30                35

gtt gtc cat gcc gac tct ggg aag gta ttc tat gat aaa gac ata gat 259
Val Val His Ala Asp Ser Gly Lys Val Phe Tyr Asp Lys Asp Ile Asp
Val Val His Ala Asp Ser Gly Lys Val Phe Tyr Asp Lys Asp Ile Asp
                        40                45                50

gct gta atc tat cct gcc agc atg acg aaa atc gca act gcc ctc ttt 307
Ala Val Ile Tyr Pro Ala Ser Met Thr Lys Ile Ala Thr Ala Leu Phe
Ala Val Ile Tyr Pro Ala Ser Met Thr Lys Ile Ala Thr Ala Leu Phe
                        55                60                65

atc cta aag cac tat ccc aca gtc ctc gat act ctc atc aaa gtc aaa 355
Ile Leu Lys His Tyr Pro Thr Val Leu Asp Thr Leu Ile Lys Val Lys
Ile Leu Lys His Tyr Pro Thr Val Leu Asp Thr Leu Ile Lys Val Lys
                        70                75                80                85

caa gat gcg atc gct tcc atc act ccg caa gca aaa aaa caa tca gga 403
Gln Asp Ala Ile Ala Ser Ile Thr Pro Gln Ala Lys Lys Gln Ser Gly
Gln Asp Ala Ile Ala Ser Ile Thr Pro Gln Ala Lys Lys Gln Ser Gly
                        90                95                100

tat cgt agt cct ccc cac tgg tta gaa act gat gga tct aca ata cag 451
Tyr Arg Ser Pro Pro His Trp Leu Glu Thr Asp Gly Ser Thr Ile Gln
Tyr Arg Ser Pro Pro His Trp Leu Glu Thr Asp Gly Ser Thr Ile Gln
                        105                110                115

ctc cat ctt cga gaa gag ctt tta ggg tgg gac ctg ttc cac gcc tta 499
Leu His Leu Arg Glu Glu Leu Leu Gly Trp Asp Leu Phe His Ala Leu
Leu His Leu Arg Glu Glu Leu Leu Gly Trp Asp Leu Phe His Ala Leu
                        120                125                130

ctg gtc tgt tct gct aat gat gct gcg aat gtc tta gct atg gca tgt 547
Leu Val Cys Ser Ala Asn Asp Ala Ala Asn Val Leu Ala Met Ala Cys
Leu Val Cys Ser Ala Asn Asp Ala Ala Asn Val Leu Ala Met Ala Cys
                        135                140                145

tgc gga tct gta gag aag ttt atg gat aag ctg aac ttc ttc tta aaa 595
Cys Gly Ser Val Glu Lys Phe Met Asp Lys Leu Asn Phe Phe Leu Lys
Cys Gly Ser Val Glu Lys Phe Met Asp Lys Leu Asn Phe Phe Leu Lys
                        150                155                160                165

gaa gaa atc ggc tgc act cat acc cat ttt aat aat ccc cat ggg tta 643
Glu Glu Ile Gly Cys Thr His Thr His Phe Asn Asn Pro His Gly Leu
Glu Glu Ile Gly Cys Thr His Thr His Phe Asn Asn Pro His Gly Leu
                        170                175                180

```

Fig. 11 (con't)

cat	cat	cog	aat	cac	tat	act	aca	acc	cgt	gat	ctt	att	agc	atc	atg	691
His	His	Pro	Asn	His	Tyr	Thr	Thr	Thr	Arg	Asp	Leu	Ile	Ser	Ile	Met	
His	His	Pro	Asn	His	Tyr	Thr	Thr	Thr	Arg	Asp	Leu	Ile	Ser	Ile	Met	
			185						190					195		
cgt	tgc	gct	ctg	aaa	gaa	cct	cca	ttt	cga	ggg	gtc	atc	tcc	acg	aca	739
Arg	Cys	Ala	Leu	Lys	Glu	Pro	Pro	Phe	Arg	Gly	Val	Ile	Ser	Thr	Thr	
Arg	Cys	Ala	Leu	Lys	Glu	Pro	Pro	Phe	Arg	Gly	Val	Ile	Ser	Thr	Thr	
			200					205				210				
agc	tat	aaa	ata	ggg	gct	aca	aac	ctg	cat	ggc	gaa	cgg	atc	cta	tcc	787
Ser	Tyr	Lys	Ile	Gly	Ala	Thr	Asn	Leu	His	Gly	Glu	Arg	Ile	Leu	Ser	
Ser	Tyr	Lys	Ile	Gly	Ala	Thr	Asn	Leu	His	Gly	Glu	Arg	Ile	Leu	Ser	
			215			220					225					
cca	aca	aac	aaa	ttg	ctt	ctt	cct	ggg	tct	acc	tac	cac	tat	ccc	cca	835
Pro	Thr	Asn	Lys	Leu	Leu	Leu	Pro	Gly	Ser	Thr	Tyr	His	Tyr	Pro	Pro	
Pro	Thr	Asn	Lys	Leu	Leu	Leu	Pro	Gly	Ser	Thr	Tyr	His	Tyr	Pro	Pro	
					235					240					245	
gct	tta	gga	ggg	aaa	aca	ggg	acc	acc	aag	act	gca	ggg	aaa	aat	cta	883
Ala	Leu	Gly	Gly	Lys	Thr	Gly	Thr	Thr	Lys	Thr	Ala	Gly	Lys	Asn	Leu	
Ala	Leu	Gly	Gly	Lys	Thr	Gly	Thr	Thr	Lys	Thr	Ala	Gly	Lys	Asn	Leu	
				250					255					260		
att	atg	gct	gct	gaa	aaa	aat	aac	cgc	ctc	ttg	gta	acg	atc	gca	acg	931
Ile	Met	Ala	Ala	Glu	Lys	Asn	Asn	Arg	Leu	Leu	Val	Thr	Ile	Ala	Thr	
Ile	Met	Ala	Ala	Glu	Lys	Asn	Asn	Arg	Leu	Leu	Val	Thr	Ile	Ala	Thr	
				265				270					275			
ggc	tat	tgc	ggt	cct	gtg	agt	gat	ctc	tac	caa	gat	gtc	att	gct	cta	979
Gly	Tyr	Ser	Gly	Pro	Val	Ser	Asp	Leu	Tyr	Gln	Asp	Val	Ile	Ala	Leu	
Gly	Tyr	Ser	Gly	Pro												
			280					285				290				
tgt	gaa	acg	gta	ttt	aac	gag	ccg	cta	tta	aga	aaa	gag	ctc	gtc	ccc	1027
Cys	Glu	Thr	Val	Phe	Asn	Glu	Pro	Leu	Leu	Arg	Lys	Glu	Leu	Val	Pro	
			295			300					305					
ccc	tcc	gac	tgt	ctc	caa	tta	gaa	ata	gcg	aat	ctt	ggg	aag	ctt	tct	1075
Pro	Ser	Asp	Cys	Leu	Gln	Leu	Glu	Ile	Ala	Asn	Leu	Gly	Lys	Leu	Ser	
					315					320					325	
tgc	cct	ctt	cct	gag	gga	ctc	tac	tat	gac	ttc	tat	gcc	tcc	gaa	gat	1123
Cys	Pro	Leu	Pro	Glu	Gly	Leu	Tyr	Tyr	Asp	Phe	Tyr	Ala	Ser	Glu	Asp	
				330					335					340		
cgc	gaa	cct	ctt	tct	gta	tct	ttt	att	gca	cat	gcg	gac	gcc	ttc	cct	1171
Arg	Glu	Pro	Leu	Ser	Val	Ser	Phe	Ile	Ala	His	Ala	Asp	Ala	Phe	Pro	
				345				350					355			
att	gaa	caa	gga	gat	ctt	ctt	ggt	cat	tgg	gtt	ttt	tat	gac	gat	gaa	1219
Ile	Glu	Gln	Gly	Asp	Leu	Leu	Gly	His	Trp	Val	Phe	Tyr	Asp	Asp	Glu	
			360				365					370				

Fig. 11 (con't)

```

ggc aag aaa att tct tcc cag cct ttc tat gcc cct tgt cgt ttt gag 1267
Gly Lys Lys Ile Ser Ser Gln Pro Phe Tyr Ala Pro Cys Arg Phe Glu
    375                      380                      385

cgc act atc aag cct tgg aaa ctc tat atg aaa cgt gtc ttc aca tcg 1315
Arg Thr Ile Lys Pro Trp Lys Leu Tyr Met Lys Arg Val Phe Thr Ser
    390                      395                      400                      405

tat aga acc tat atg tct ata acc atg ctg ctc atg tat ttt cgc atc 1363
Tyr Arg Thr Tyr Met Ser Ile Thr Met Leu Leu Met Tyr Phe Arg Ile
                410                      415                      420

cgc aag cac cgc aag tat aaa aat tta aaa cac tat tct aaa atc 1408
Arg Lys His Arg Lys Tyr Lys Asn Leu Lys His Tyr Ser Lys Ile
                425                      430                      435

taacttttttc ttttaattta taaaaaacca aagggtttatg taagatttgc gctttttcaat 1468

ccaacaagaa tcccttgtgc gcacattact tt 1500

```



Figure 12 (RY-39)

## Restriction enzyme analysis of CPN100508

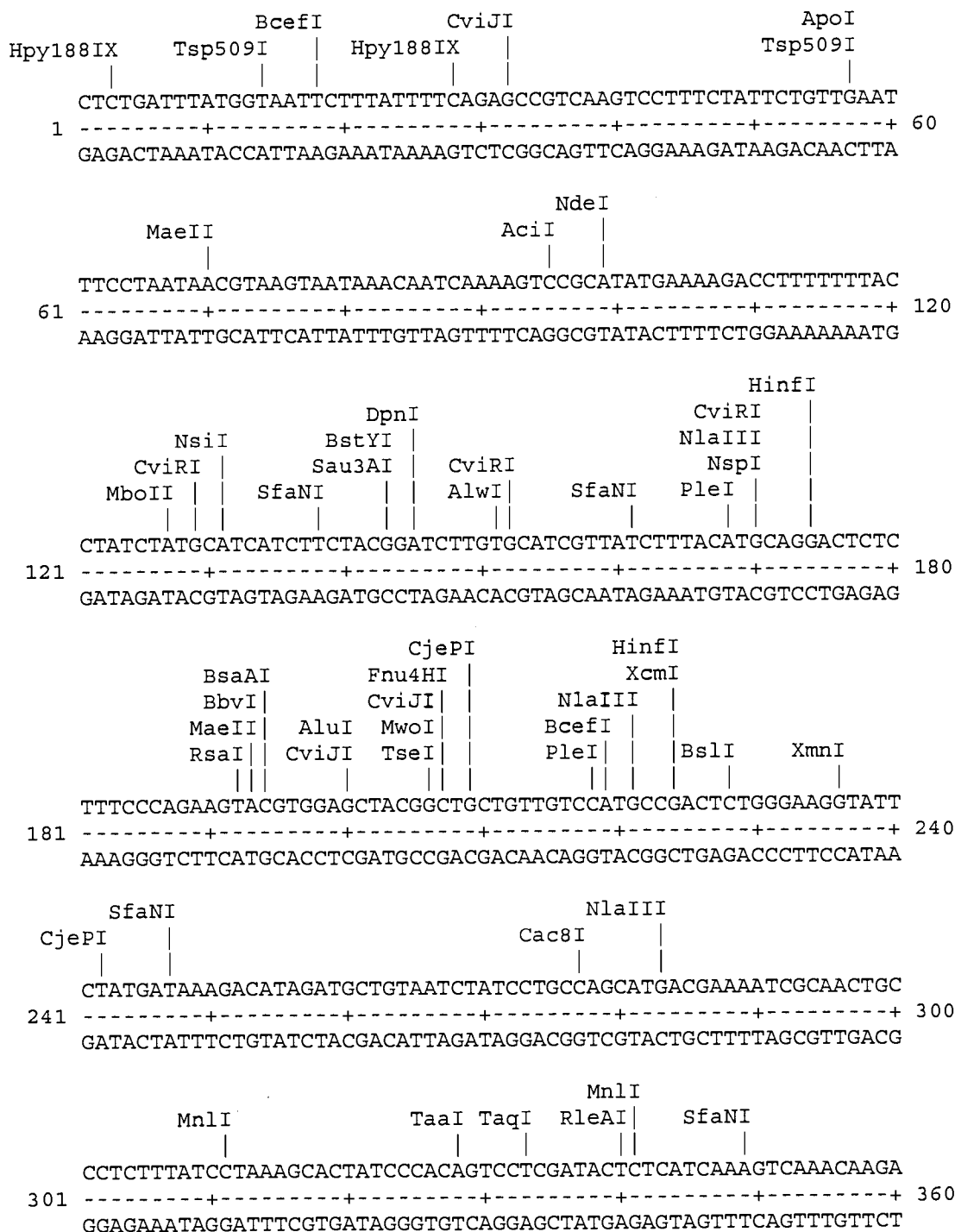


Fig. 12 (con't)

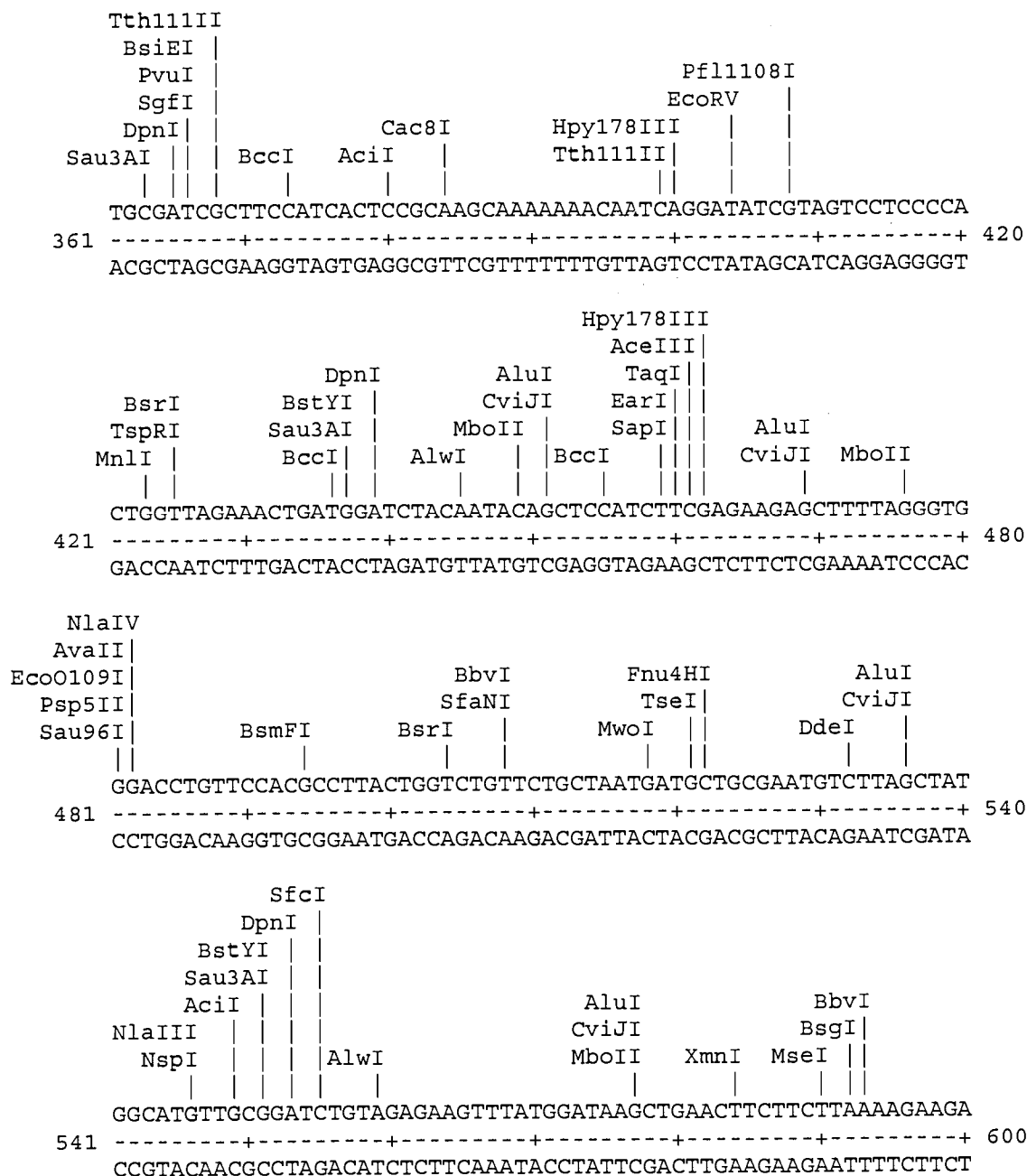


Fig. 12 (con't)

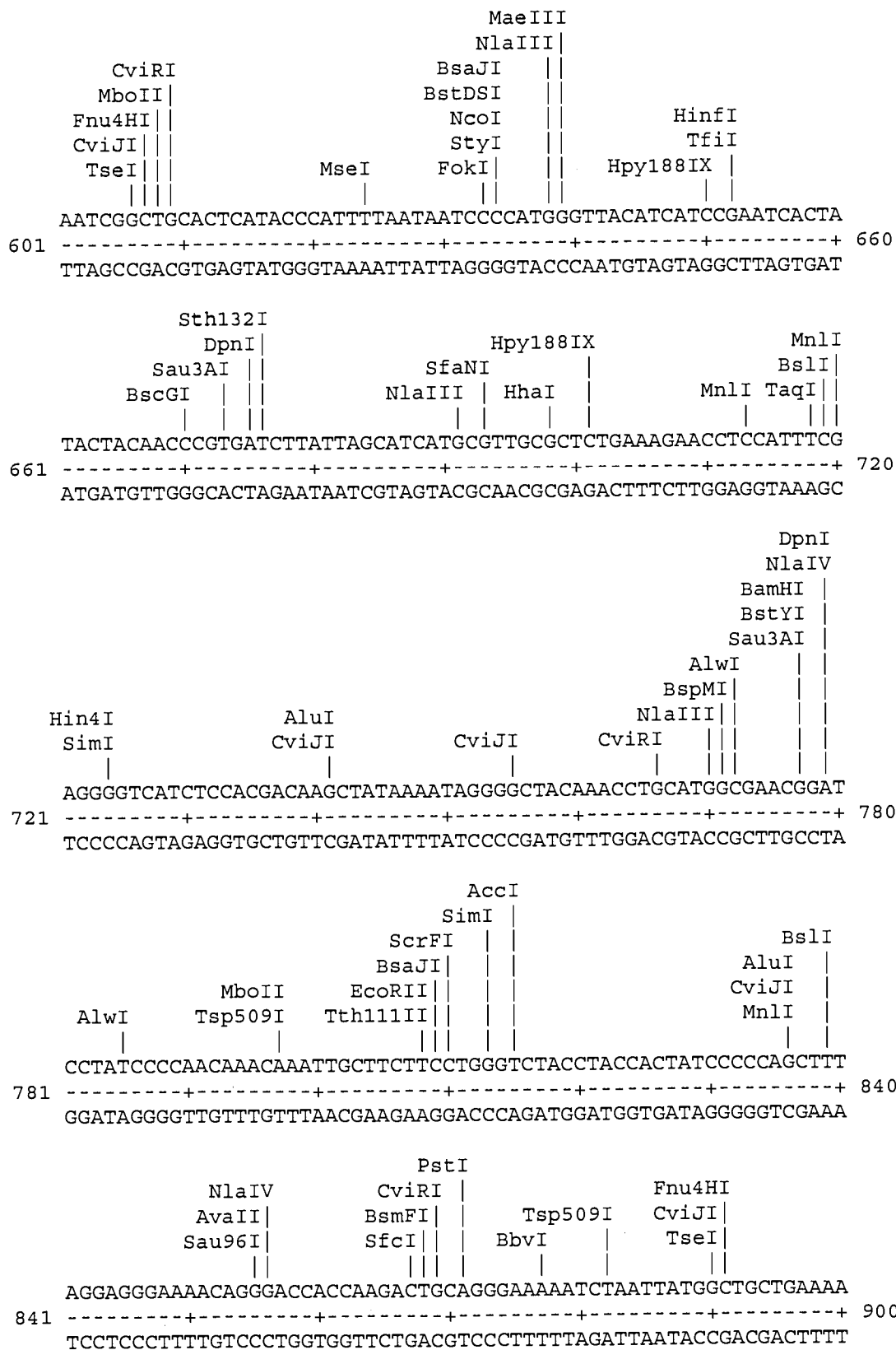


Fig. 12 (con't)

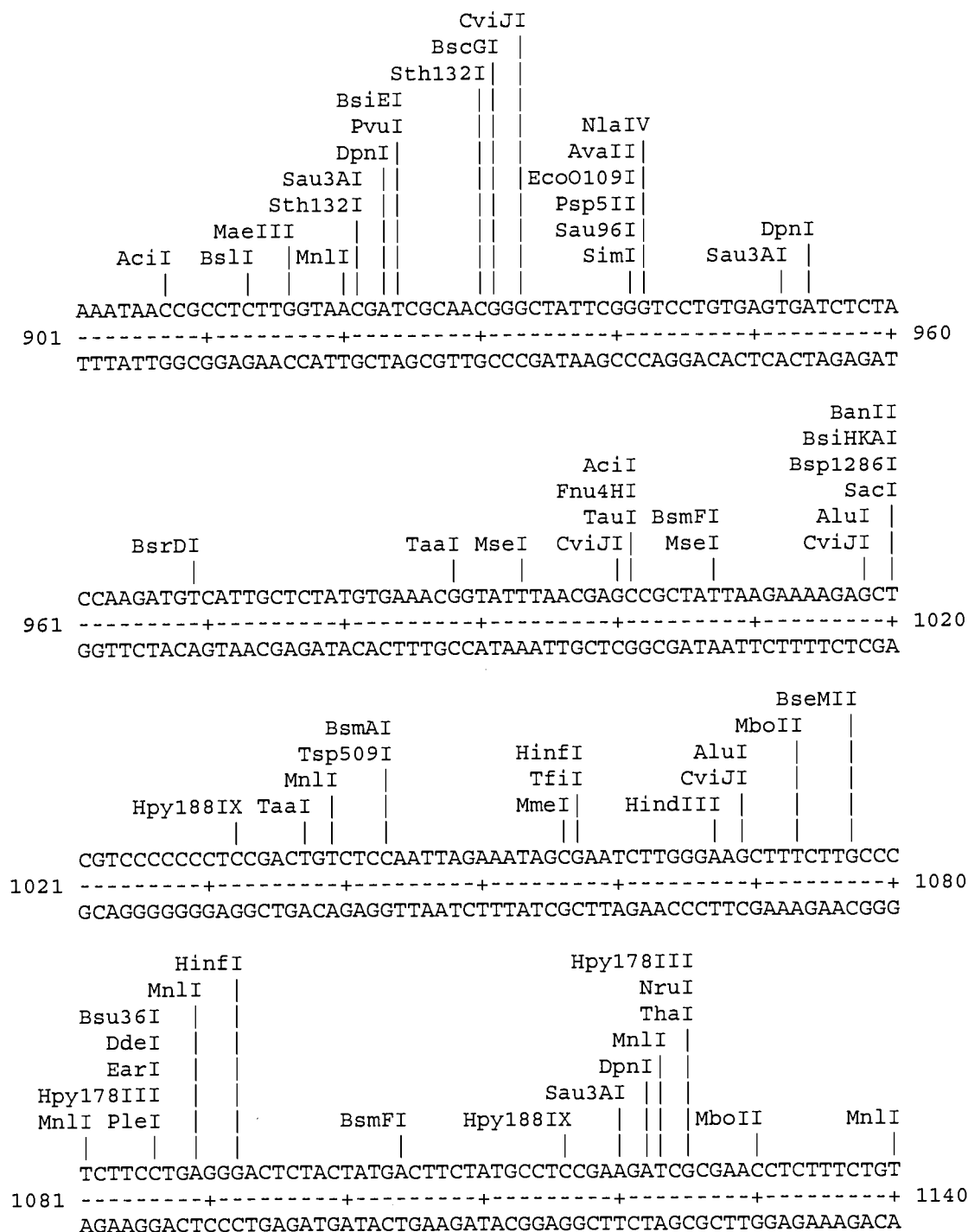


Fig. 12 (con't)

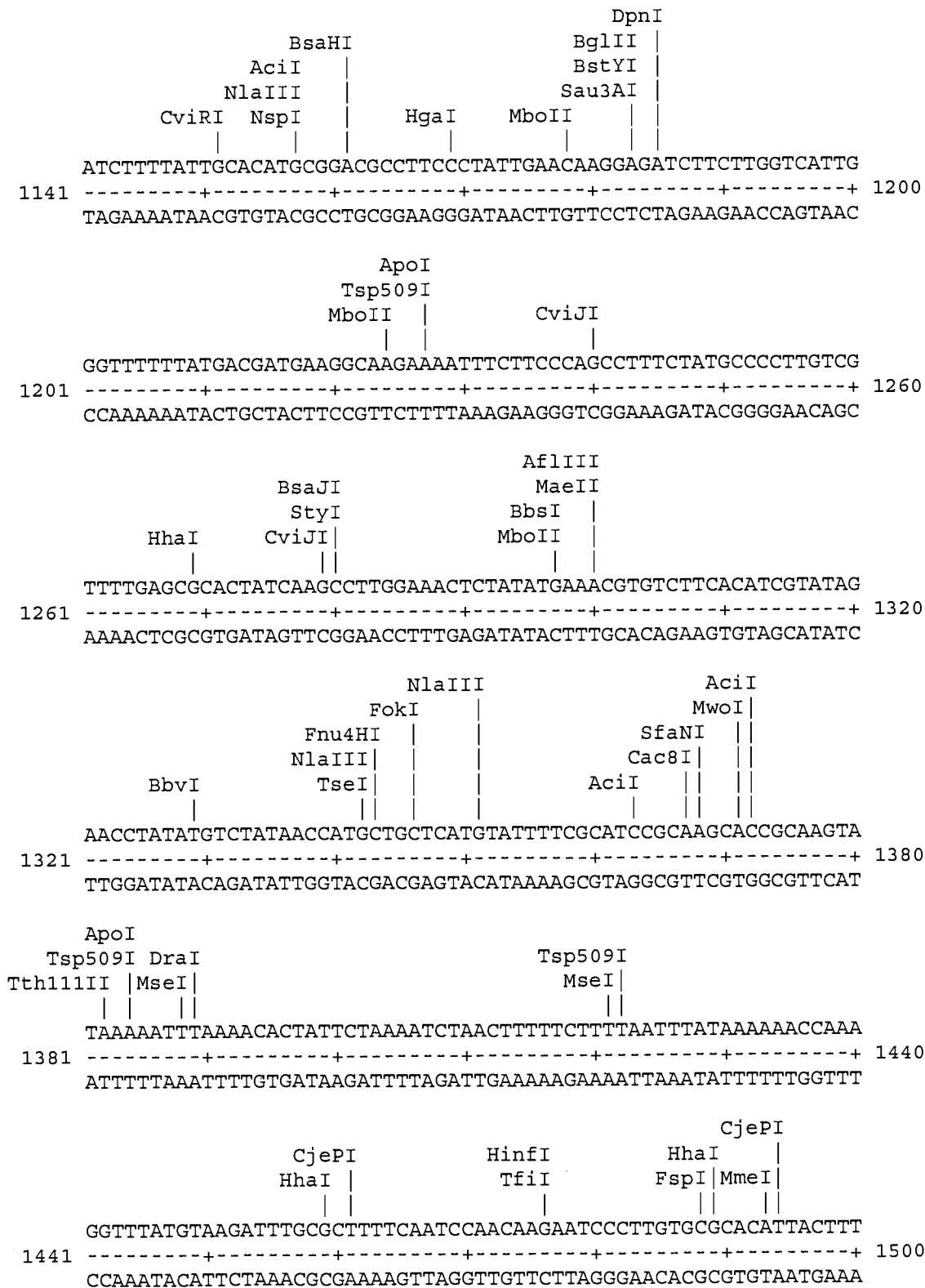


Figure 13: CPN100515-

```

aaggagcaaaa tggagattgg ccaaataagac gagcaagggt ttgcataaga atagcctttt 60
tcgcaataat aacttgccta aacgatcttg taaacgactt atg gct tct aat ccc 115
                                         Met Ala Ser Asn Pro
                                         1           5

att tta cag ata gag gat cta tcc ata acc ttg gca aaa caa cgc caa 163
Ile Leu Gln Ile Glu Asp Leu Ser Ile Thr Leu Ala Lys Gln Arg Gln
Ile Leu Gln Ile Glu Asp Leu Ser Ile Thr Leu Ala Lys Gln Arg Gln
                               10                15                20

cag tac ccc atc gtc caa tct tta tct ttt act atc aat gaa gga caa 211
Gln Tyr Pro Ile Val Gln Ser Leu Ser Phe Thr Ile Asn Glu Gly Gln
Gln Tyr Pro Ile Val Gln Ser Leu Ser Phe Thr Ile Asn Glu Gly Gln
                               25                30                35

acc tta gca atc att gga gaa tca gga tca gga aaa tct gtc tct gcg 259
Thr Leu Ala Ile Ile Gly Glu Ser Gly Ser Gly Lys Ser Val Ser Ala
Thr Leu Ala Ile Ile Gly Glu Ser Gly Ser Gly Lys Ser Val Ser Ala
                               40                45                50

cat gca atc ctt cga tta ctt cct tgc ccc cca ttt tct gtt tct ggc 307
His Ala Ile Leu Arg Leu Leu Pro Cys Pro Pro Phe Ser Val Ser Gly
His Ala Ile Leu Arg Leu Leu Pro Cys Pro Pro Phe Ser Val Ser Gly
                               55                60                65

cag gtc aac ttc caa ggc cac aac tta ctt acg gct tct cgc tct ata 355
Gln Val Asn Phe Gln Gly His Asn Leu Leu Thr Ala Ser Arg Ser Ile
Gln Val Asn Phe Gln Gly His Asn Leu Leu Thr Ala Ser Arg Ser Ile
                               70                75                80                85

caa aaa aag att ata ggg aca gaa att tct atg atc ttt caa aac ccg 403
Gln Lys Lys Ile Ile Gly Thr Glu Ile Ser Met Ile Phe Gln Asn Pro
Gln Lys Lys Ile Ile Gly Thr Glu Ile Ser Met Ile Phe Gln Asn Pro
                               90                95                100

caa gca tct cta aac ccc gtg ttt act att gaa cag cag ttt cga gaa 451
Gln Ala Ser Leu Asn Pro Val Phe Thr Ile Glu Gln Gln Phe Arg Glu
Gln Ala Ser Leu Asn Pro Val Phe Thr Ile Glu Gln Gln Phe Arg Glu
                               105                110                115

att att cat acc cac cta gcc tta act gca gaa gtt gct aaa gaa aag 499
Ile Ile His Thr His Leu Ala Leu Thr Ala Glu Val Ala Lys Glu Lys
Ile Ile His Thr His Leu Ala Leu Thr Ala Glu Val Ala Lys Glu Lys
                               120                125                130

atg tta tac gct ctt gaa gaa aca ggg ttt cat gat ccc agg ctg tgc 547
Met Leu Tyr Ala Leu Glu Glu Thr Gly Phe His Asp Pro Arg Leu Cys
Met Leu Tyr Ala Leu Glu Glu Thr Gly Phe His Asp Pro Arg Leu Cys
                               135                140                145

ttg aat ctc tac ccc cac caa ctc tct gga ggg atg ctt caa aga att 595
Leu Asn Leu Tyr Pro His Gln Leu Ser Gly Gly Met Leu Gln Arg Ile
Leu Asn Leu Tyr Pro His Gln Leu Ser Gly Gly Met Leu Gln Arg Ile
                               150                155                160                165

```

Fig. 13 (con't)

tgc att gcc atg gcg ctc ctc tgt tct cct aaa ctt ctt att gct gat	643
Cys Ile Ala Met Ala Leu Leu Cys Ser Pro Lys Leu Leu Ile Ala Asp	
Cys Ile Ala Met Ala Leu Leu Cys Ser Pro Lys Leu Leu Ile Ala Asp	
170 175 180	
gaa cct acg act gct tta gat gtt tct gtt cag tat cag att cta caa	691
Glu Pro Thr Thr Ala Leu Asp Val Ser Val Gln Tyr Gln Ile Leu Gln	
Glu Pro Thr Thr Ala Leu Asp Val Ser Val Gln Tyr Gln Ile Leu Gln	
185 190 195	
tta cta aaa aca cta cag aaa aaa acg gga atg agc ctt ctt att att	739
Leu Leu Lys Thr Leu Gln Lys Lys Thr Gly Met Ser Leu Leu Ile Ile	
Leu Leu Lys Thr Leu Gln Lys Lys Thr Gly Met Ser Leu Leu Ile Ile	
200 205 210	
acc cat aat atg gga gtc gtt gca gaa act gct gat gac gtg ctc gtg	787
Thr His Asn Met Gly Val Val Ala Glu Thr Ala Asp Asp Val Leu Val	
Thr His Asn Met Gly Val Val Ala Glu Thr Ala Asp Asp Val Leu Val	
215 220 225	
ctc tat gca gga cgc atg gta gaa tgt gcc cct gcg gtt caa atg ttc	835
Leu Tyr Ala Gly Arg Met Val Glu Cys Ala Pro Ala Val Gln Met Phe	
Leu Tyr Ala Gly Arg Met Val Glu Cys Ala Pro Ala Val Gln Met Phe	
230 235 240 245	
cat aat cct tct cat ccc tat acc cga gat ctt tta gca tcc aga ccc	883
His Asn Pro Ser His Pro Tyr Thr Arg Asp Leu Leu Ala Ser Arg Pro	
His Asn Pro Ser His Pro Tyr Thr Arg Asp Leu Leu Ala Ser Arg Pro	
250 255 260	
tct cta caa ccg caa caa cta ggt tcc ttc aac ccc att cca gga cag	931
Ser Leu Gln Pro Gln Gln Leu Gly Ser Phe Asn Pro Ile Pro Gly Gln	
Ser Leu Gln Pro Gln Gln Leu Gly Ser Phe Asn Pro Ile Pro Gly Gln	
265 270 275	
ccc cca cac tac acg gcc ttt ccc tcg gga tgt cgc tat cac cct aga	979
Pro Pro His Tyr Thr Ala Phe Pro Ser Gly Cys Arg Tyr His Pro Arg	
Pro Pro His Tyr Thr Ala Phe Pro Ser Gly Cys Arg Tyr His Pro Arg	
280 285 290	
tgc tca aaa att tta aat cga tgt tct gcg gaa gct cca gaa atc tat	1027
Cys Ser Lys Ile Leu Asn Arg Cys Ser Ala Glu Ala Pro Glu Ile Tyr	
Cys Ser Lys Ile Leu Asn Arg Cys Ser Ala Glu Ala Pro Glu Ile Tyr	
295 300 305	
ccg gta cgc gaa ggt cac aaa gta agg gtt gcc tgt atg acg act aat	1075
Pro Val Arg Glu Gly His Lys Val Arg Val Gly Cys Met Thr Thr Asn	
Pro Val Arg Glu Gly His Lys Val Arg Val Gly Cys Met Thr Thr Asn	
310 315 320 325	
ttt ccc caa cct tta att caa gca acc tca tta aca aag cac tat tac	1123
Phe Pro Gln Pro Leu Ile Gln Ala Thr Ser Leu Thr Lys His Tyr Tyr	
Phe Pro Gln Pro Leu Ile Gln Ala Thr Ser Leu Thr Lys His Tyr Tyr	
330 335 340	

Fig. 13 (con't)

aag cgt tcc ttt tgg ttt cag gga aag aca att gcc agt cgt cct gtt	1171
Lys Arg Ser Phe Trp Phe Gln Gly Lys Thr Ile Ala Ser Arg Pro Val	
Lys Arg Ser Phe Trp Phe Gln Gly Lys Thr Ile Ala Ser Arg Pro Val	
345 350 355	
gac gac gtc tct ttt tca cta tac tcc aga cgt gct gtc gga ctt att	1219
Asp Asp Val Ser Phe Ser Leu Tyr Ser Arg Arg Ala Val Gly Leu Ile	
Asp Asp Val Ser Phe Ser Leu Tyr Ser Arg Arg Ala Val Gly Leu Ile	
360 365 370	
gga gaa tct gga tca ggg aaa agt acc ctg gcg tta gct ctc gca ggt	1267
Gly Glu Ser Gly Ser Gly Lys Ser Thr Leu Ala Leu Ala Leu Ala Gly	
Gly Glu Ser Gly Ser Gly Lys Ser Thr Leu Ala Leu Ala Leu Ala Gly	
375 380 385	
ctc cta cct ctc acc tct ggg ttc tta act ttt aac ggc acc cca atc	1315
Leu Leu Pro Leu Thr Ser Gly Phe Leu Thr Phe Asn Gly Thr Pro Ile	
Leu Leu Pro Leu Thr Ser Gly Phe Leu Thr Phe Asn Gly Thr Pro Ile	
390 395 400 405	
aag ttg cat tct aaa cac gga cgc cat caa tta cga tct caa gta cgg	1363
Lys Leu His Ser Lys His Gly Arg His Gln Leu Arg Ser Gln Val Arg	
Lys Leu His Ser Lys His Gly Arg His Gln Leu Arg Ser Gln Val Arg	
410 415 420	
ttg gtc ttt caa aat cca caa gct tca tta aac ccg cga aaa act atc	1411
Leu Val Phe Gln Asn Pro Gln Ala Ser Leu Asn Pro Arg Lys Thr Ile	
Leu Val Phe Gln Asn Pro Gln Ala Ser Leu Asn Pro Arg Lys Thr Ile	
425 430 435	
cta gat agt tta ggc cac tct ctg ctt tac cat aaa ctc gtc cca aaa	1459
Leu Asp Ser Leu Gly His Ser Leu Leu Tyr His Lys Leu Val Pro Lys	
Leu Asp Ser Leu Gly His Ser Leu Leu Tyr His Lys Leu Val Pro Lys	
440 445 450	
gaa aaa gta cta gca acg gta agg gaa tat tta gaa ttg gta ggg tta	1507
Glu Lys Val Leu Ala Thr Val Arg Glu Tyr Leu Glu Leu Val Gly Leu	
Glu Lys Val Leu Ala Thr Val Arg Glu Tyr Leu Glu Leu Val Gly Leu	
455 460 465	
tct gag gag tat ttt tat cgt tat cct cac cag ctt tct gga gga caa	1555
Ser Glu Glu Tyr Phe Tyr Arg Tyr Pro His Gln Leu Ser Gly Gly Gln	
Ser Glu Glu Tyr Phe Tyr Arg Tyr Pro His Gln Leu Ser Gly Gly Gln	
470 475 480 485	
caa caa cga gtc tct ata gcg aga gcc cta tta gga gtc cct cag tta	1603
Gln Gln Arg Val Ser Ile Ala Arg Ala Leu Leu Gly Val Pro Gln Leu	
Gln Gln Arg Val Ser Ile Ala Arg Ala Leu Leu Gly Val Pro Gln Leu	
490 495 500	
att att tgt gac gaa att gtt tct gct cta gat tta tct att caa gca	1651
Ile Ile Cys Asp Glu Ile Val Ser Ala Leu Asp Leu Ser Ile Gln Ala	
Ile Ile Cys Asp Glu Ile Val Ser Ala Leu Asp Leu Ser Ile Gln Ala	
505 510 515	
caa att ctg aat atg ctt gcc gag ctg caa aaa aaa ctc agc ctc aca	1699
Gln Ile Leu Asn Met Leu Ala Glu Leu Gln Lys Lys Leu Ser Leu Thr	
Gln Ile Leu Asn Met Leu Ala Glu Leu Gln Lys Lys Leu Ser Leu Thr	
520 525 530	



Fig. 13 (con't)

tat ctc ttc att tcg cat gat ctt gcc gtt gta cgc tcg ttc tgc aca	1747
Tyr Leu Phe Ile Ser His Asp Leu Ala Val Val Arg Ser Phe Cys Thr	
Tyr Leu Phe Ile Ser His Asp Leu Ala Val Val Arg Ser Phe Cys Thr	
535 540 545	
gag gta ttc att atg tat aag ggg caa att gta gaa aaa gga aat aca	1795
Glu Val Phe Ile Met Tyr Lys Gly Gln Ile Val Glu Lys Gly Asn Thr	
Glu Val Phe Ile Met Tyr Lys Gly Gln Ile Val Glu Lys Gly Asn Thr	
550 555 560 565	
aaa cgc att ttt tct gat cca caa cat cct tat acg cgc atg ttg tta	1843
Lys Arg Ile Phe Ser Asp Pro Gln His Pro Tyr Thr Arg Met Leu Leu	
Lys Arg Ile Phe Ser Asp Pro Gln His Pro Tyr Thr Arg Met Leu Leu	
570 575 580	
aat gcc caa ctt cca gag act cct gat caa agg caa tct aaa cct ata	1891
Asn Ala Gln Leu Pro Glu Thr Pro Asp Gln Arg Gln Ser Lys Pro Ile	
Asn Ala Gln Leu Pro Glu Thr Pro Asp Gln Arg Gln	
585 590 595	
ttc caa gaa tat cac aaa gat tct gaa gaa tct tgc tct aca gga tgc	1939
Phe Gln Glu Tyr His Lys Asp Ser Glu Glu Ser Cys Ser Thr Gly Cys	
600 605 610	
tac ttt tac aat cgt tgt cca caa aaa caa gaa gct tgc aag tca gag	1987
Tyr Phe Tyr Asn Arg Cys Pro Gln Lys Gln Glu Ala Cys Lys Ser Glu	
615 620 625	
atc atc cca aat caa gga gac gcg cac cat aca tac cgt tgt atc cat	2035
Ile Ile Pro Asn Gln Gly Asp Ala His His Thr Tyr Arg Cys Ile His	
630 635 640 645	
tgattcgctcc totacgctat totttaagcta ccattaagga atcccaaggg agaggtctgc	2095
tctat	2100

Figure 14 (RY-40)

## Restriction enzyme analysis of CPN 100515

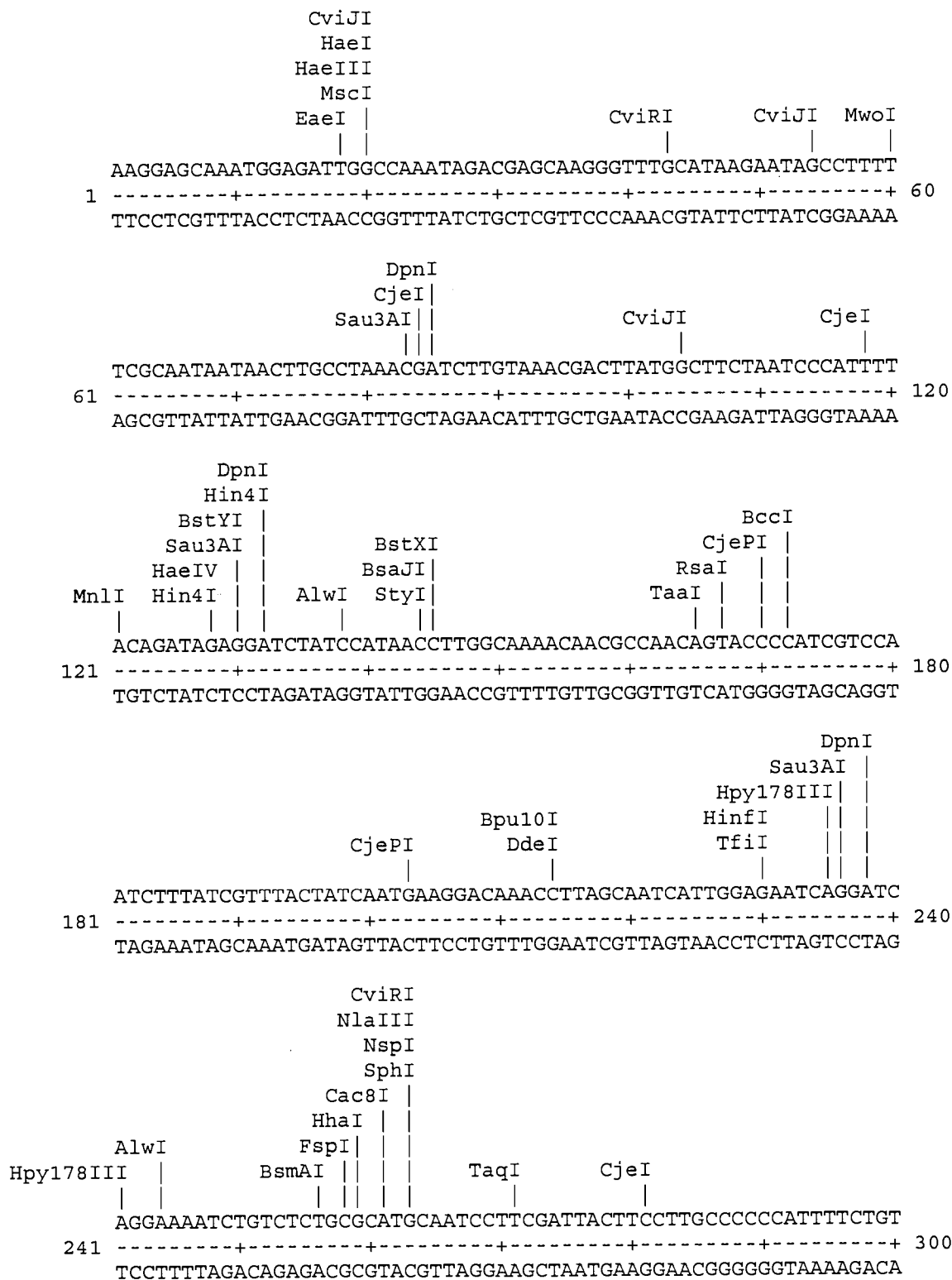


Fig. 14 (con't)

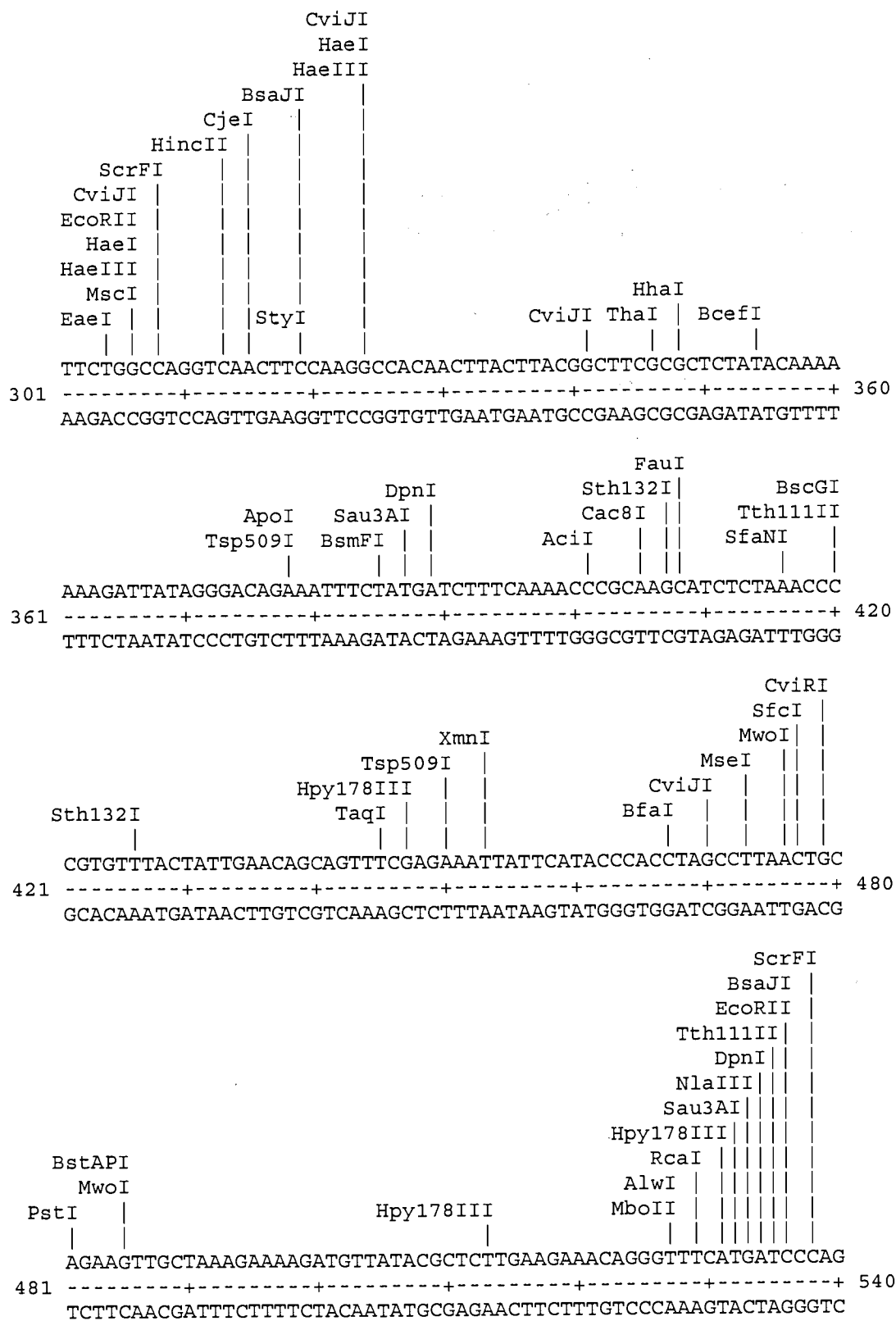


Fig. 14 (con't)

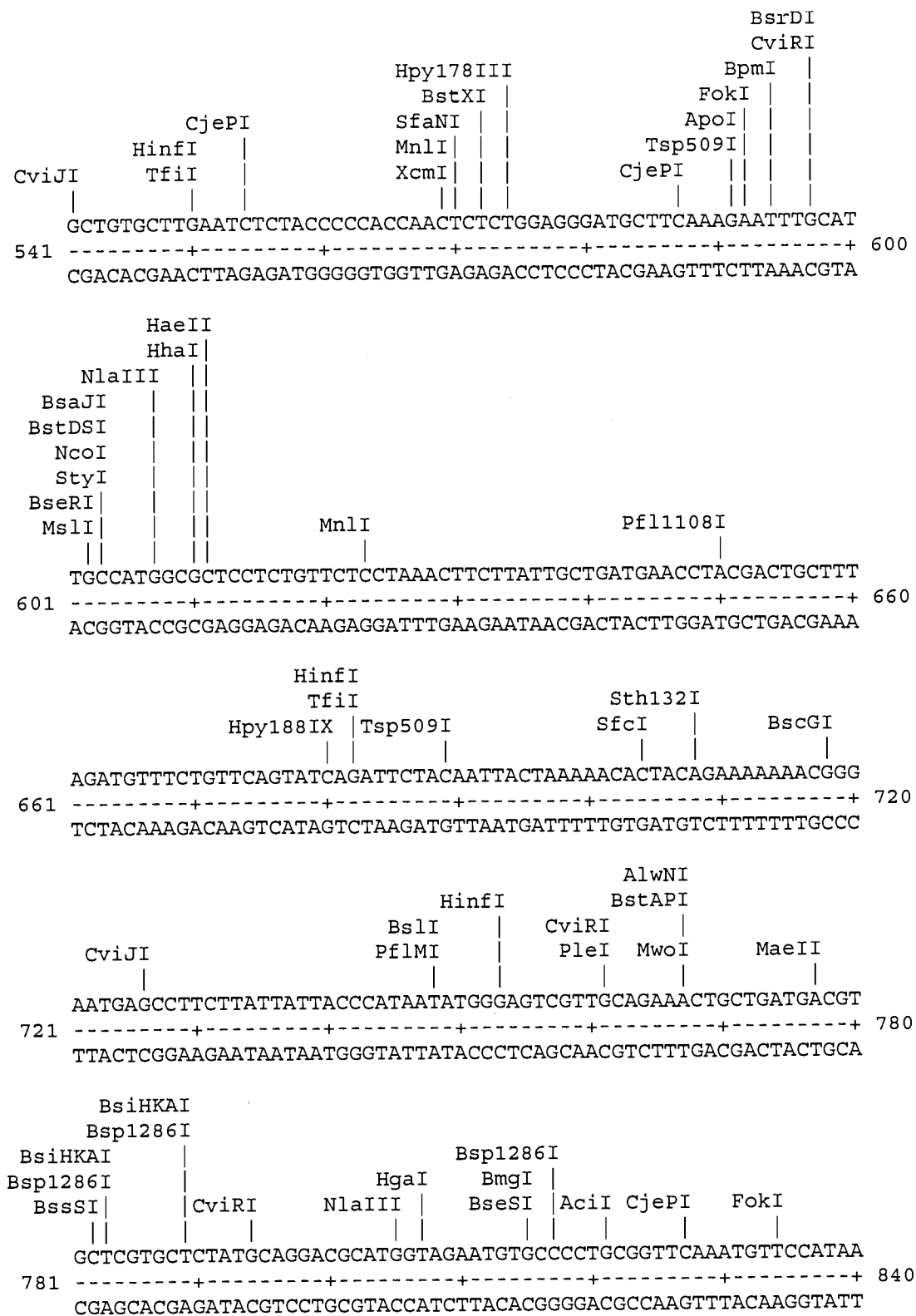


Fig. 14 (con't)

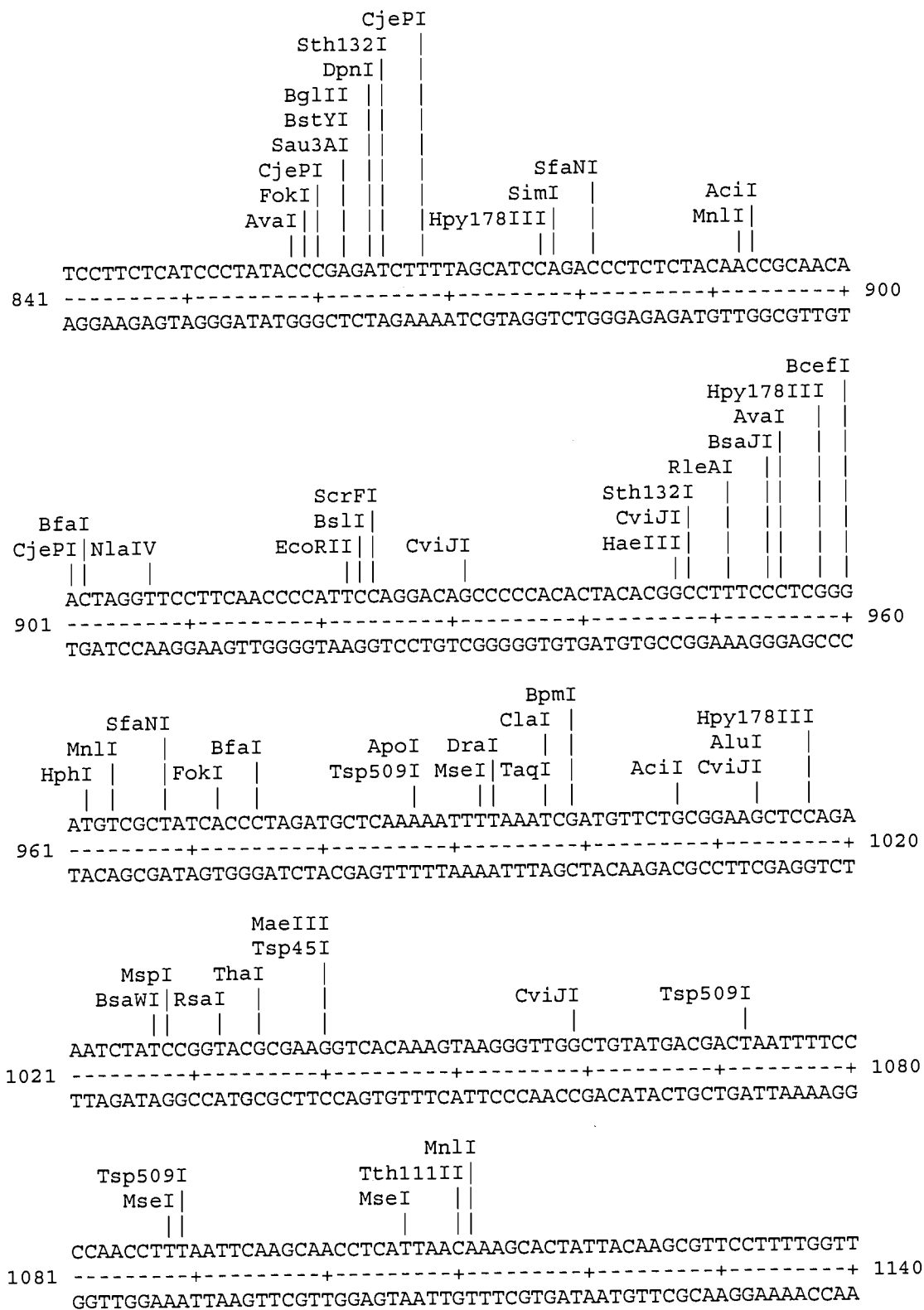


Fig. 14 (con't)

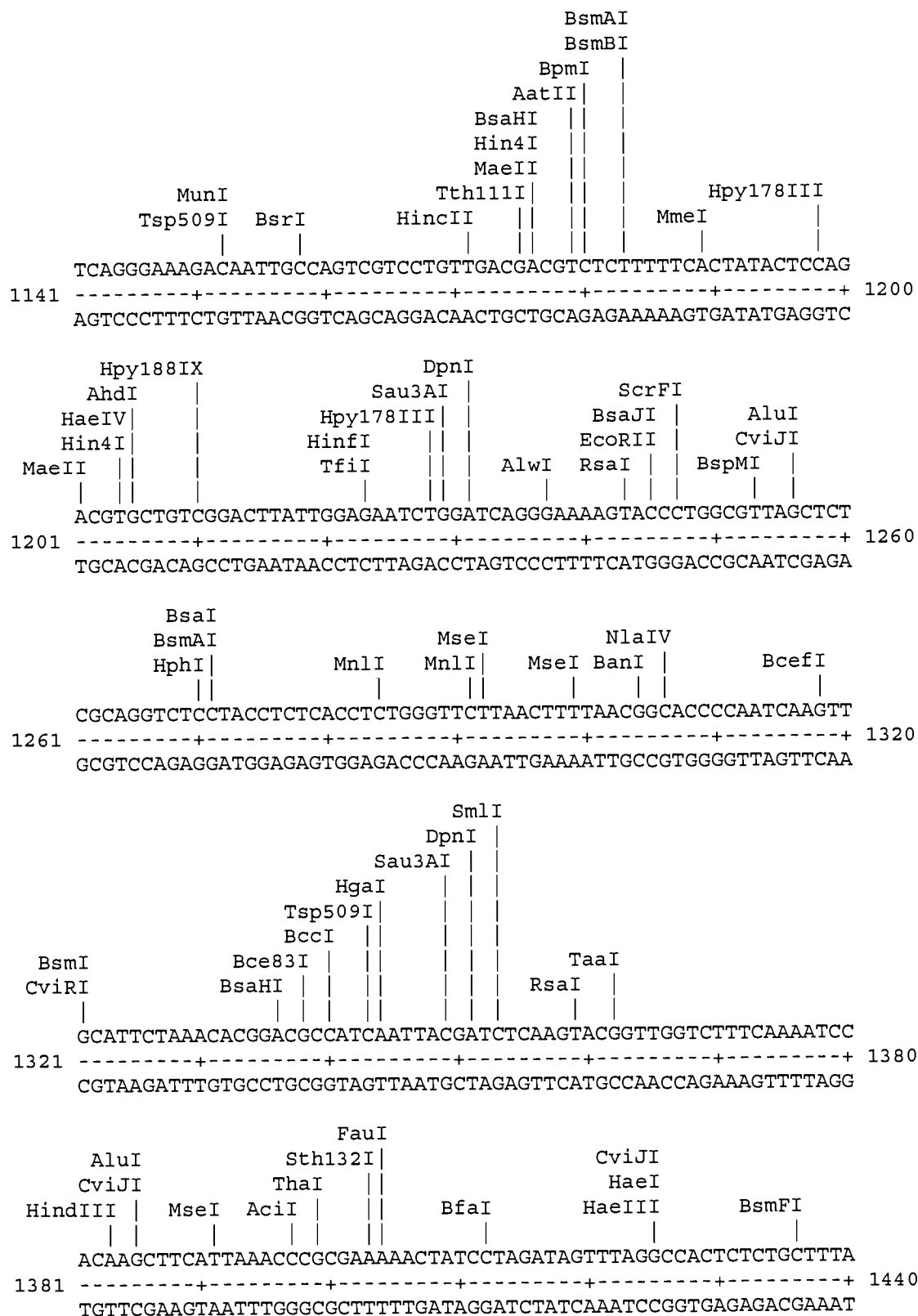


Fig. 14 (con't)

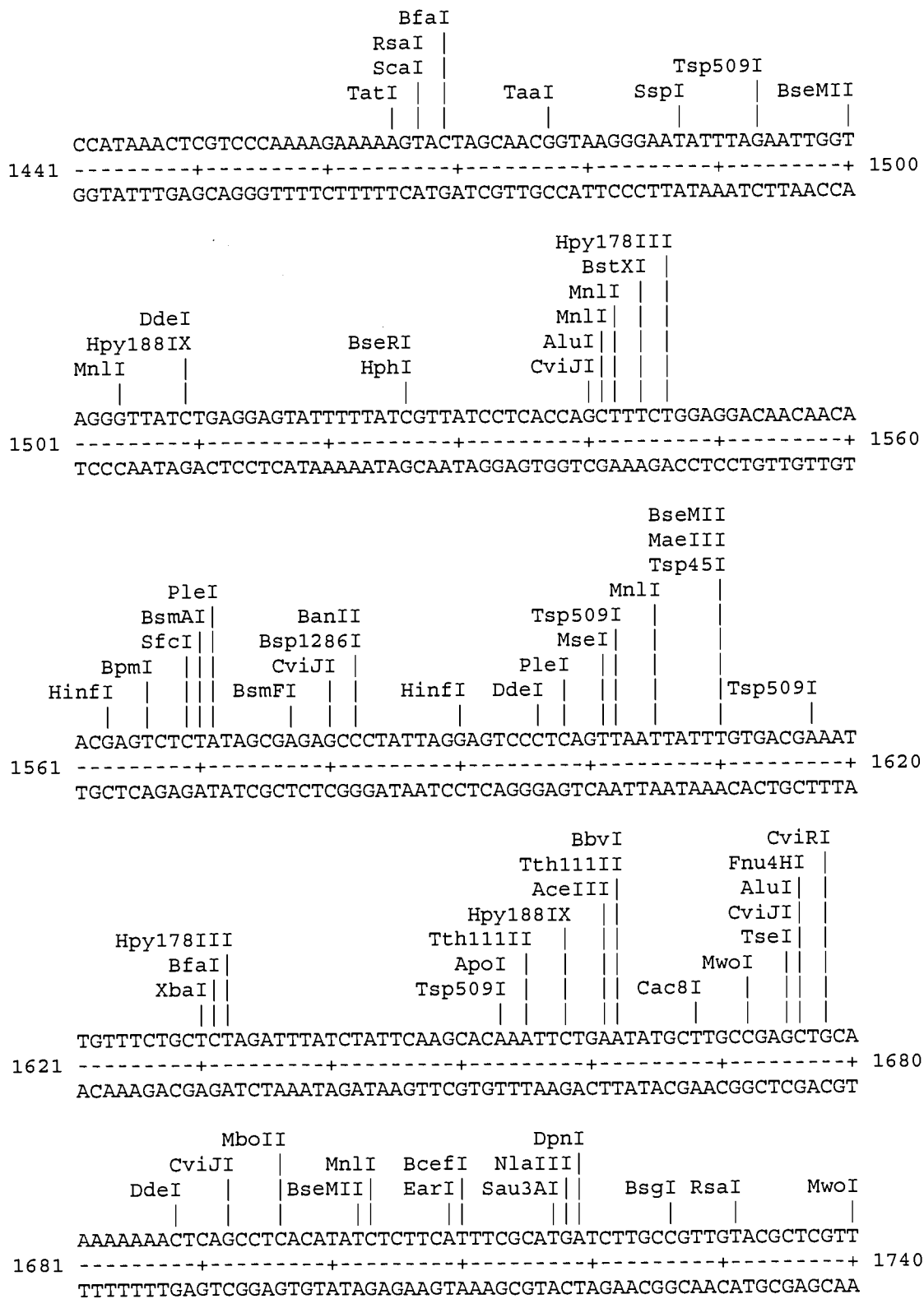


Fig. 14 (con't)

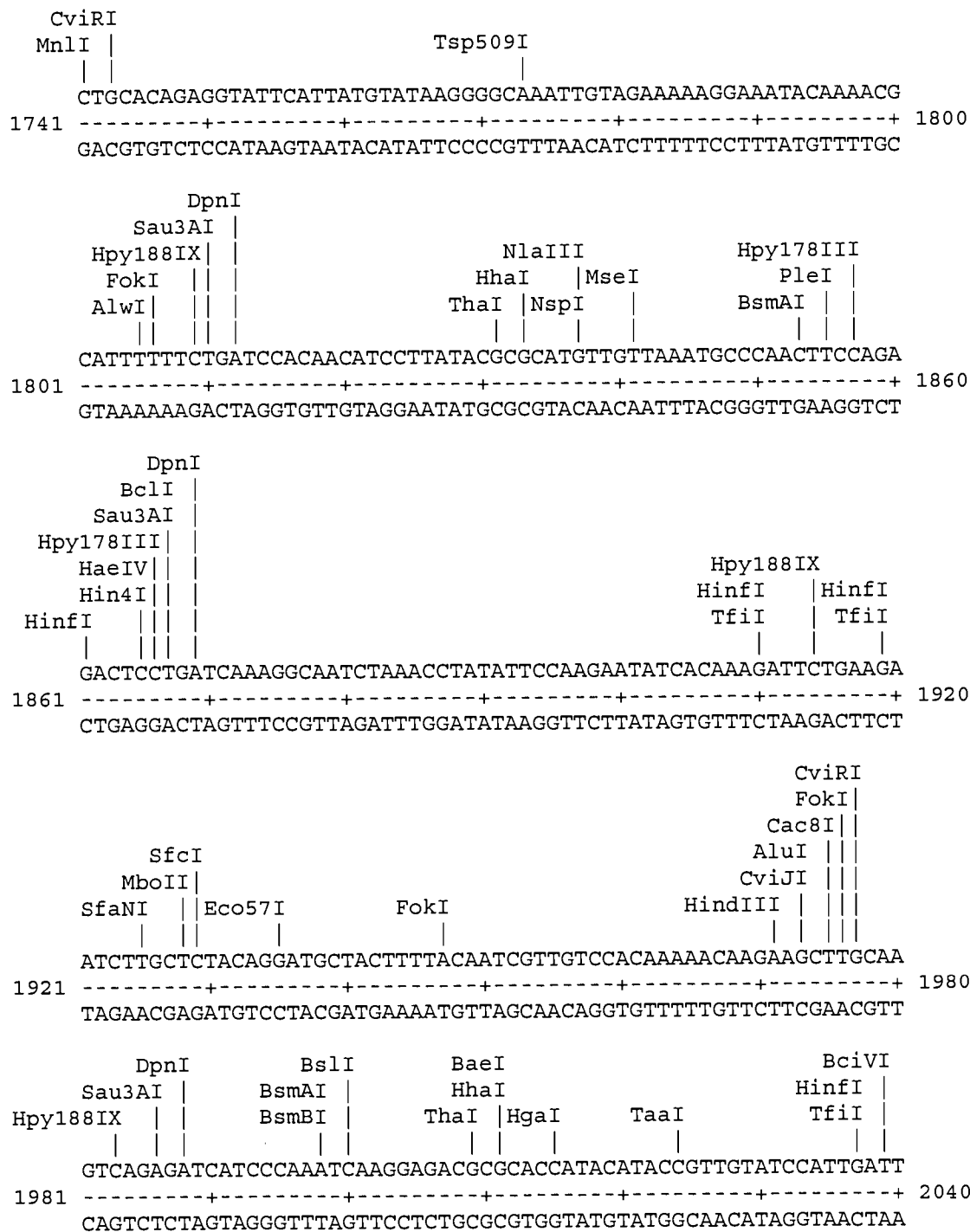




Fig. 14 (con't)



Figure 15:

```

cgaagagcaa acctccacag ttacagagaa agacgtccaa cctaaaacac aagcaacacc 60
acacgcttcg aagaaaaacg ttgcaagtcc ttcgacctct atg cca gga atc gag 115
                                         Met Pro Gly Ile Glu
                                         1           5
aaa gca gca aca aca gtg gct gta cct caa gac aaa tct gaa gaa gaa 163
Lys Ala Ala Thr Thr Val Ala Val Pro Gln Asp Lys Ser Glu Glu Glu
                        10                15                20
aaa gtt aaa gag cga ttg aca aag cgg gaa ctt acc tgt gaa gac ctt 211
Lys Val Lys Glu Arg Leu Thr Lys Arg Glu Leu Thr Cys Glu Asp Leu
                        25                30                35
aaa gat aac ggc tat act gtc aat ttt gaa gac att tct att tta gag 259
Lys Asp Asn Gly Tyr Thr Val Asn Phe Glu Asp Ile Ser Ile Leu Glu
                        40                45                50
ttg ttg cag ttc gta agt aaa att tct gga acg aac ttt gtc ttt gat 307
Leu Leu Gln Phe Val Ser Lys Ile Ser Gly Thr Asn Phe Val Phe Asp
                        55                60                65
agc aac gat ttg caa ttc aat gtc acg atc gtt tcc cac gat cct act 355
Ser Asn Asp Leu Gln Phe Asn Val Thr Ile Val Ser His Asp Pro Thr
                        70                75                80                85
tct gta gat gat tta tct aca atc tta cta caa gtc tta aaa atg cat 403
Ser Val Asp Asp Leu Ser Thr Ile Leu Leu Gln Val Leu Lys Met His
                        90                95                100
gac ttg aag gtt gtt gaa caa ggc aat aac gtc ctt atc tat cgt aat 451
Asp Leu Lys Val Val Glu Gln Gly Asn Asn Val Leu Ile Tyr Arg Asn
                        105                110                115
cct cat ctt tct aag cta tcc aca gta gtc aca gac agc tcc tta aaa 499
Pro His Leu Ser Lys Leu Ser Thr Val Val Thr Asp Ser Ser Leu Lys
                        120                125                130
gaa acg tgt gaa gct gtt gtg gtt acc cga gtg ttc cgt ctt tac agg 547
Glu Thr Cys Glu Ala Val Val Val Thr Arg Val Phe Arg Leu Tyr Arg
                        135                140                145
cgt cag ccc tct gca gca gta aat att att caa cct tta ctt tcc cat 595
Arg Gln Pro Ser Ala Ala Val Asn Ile Ile Gln Pro Leu Leu Ser His
                        150                155                160                165
gat gct atc gtt agt gct tca gaa gct act cgt cat gtt atc atc tcg 643
Asp Ala Ile Val Ser Ala Ser Glu Ala Thr Arg His Val Ile Ile Ser
                        170                175                180
gat att gct ggt aat gtc gat aaa gtc agt gat ttg cta gca gct cta 691
Asp Ile Ala Gly Asn Val Asp Lys Val Ser Asp Leu Leu Ala Ala Leu
                        185                190                195

```

Fig. 15 con't)

gat tgc cca ggc aca tct gtg gac atg act gaa tac gaa gtt aaa tat	739
Asp Cys Pro Gly Thr Ser Val Asp Met Thr Glu Tyr Glu Val Lys Tyr	
200 205 210	
gcc aat ccc gca gct ctt gtt agc tac tgc caa gat gtt ctt ggt act	787
Ala Asn Pro Ala Ala Leu Val Ser Tyr Cys Gln Asp Val Leu Gly Thr	
215 220 225	
ctg gcc gaa gat gat gct ttc caa atg ttc atc caa cct gga acg aac	835
Leu Ala Glu Asp Asp Ala Phe Gln Met Phe Ile Gln Pro Gly Thr Asn	
230 235 240 245	
aaa att ttc gtc gtc tct tca cca cgt ctt gca aat aag gca gag cag	883
Lys Ile Phe Val Val Ser Ser Pro Arg Leu Ala Asn Lys Ala Glu Gln	
250 255 260	
ctc ctg aag tcc tta gat gtc cca gaa atg gca cat acc cta gat gat	931
Leu Leu Lys Ser Leu Asp Val Pro Glu Met Ala His Thr Leu Asp Asp	
265 270 275	
cct gca agt act gcc ttg gct ttg gga gga aca gga acc acg agc cct	979
Pro Ala Ser Thr Ala Leu Ala Leu Gly Gly Thr Gly Thr Thr Ser Pro	
280 285 290	
aag agt ttg cgg ttc ttt atg tac aag ctg aag tat caa aat gga gaa	1027
Lys Ser Leu Arg Phe Phe Met Tyr Lys Leu Lys Tyr Gln Asn Gly Glu	
295 300 305	
gtg att gct aat gcc ctc caa gat atc ggt tac aat cta tat gta acc	1075
Val Ile Ala Asn Ala Leu Gln Asp Ile Gly Tyr Asn Leu Tyr Val Thr	
310 315 320 325	
aca gct atg gac gaa gat ttc att aac act ctc aat agt atc cag tgg	1123
Thr Ala Met Asp Glu Asp Phe Ile Asn Thr Leu Asn Ser Ile Gln Trp	
330 335 340	
tta gag gtc aat aac tcc ata gtt att atc gga aac caa ggg aat gtc	1171
Leu Glu Val Asn Asn Ser Ile Val Ile Ile Gly Asn Gln Gly Asn Val	
345 350 355	
gac aga gtt att ggc ctc tta aac ggt tta gat tta cct cct aaa cag	1219
Asp Arg Val Ile Gly Leu Leu Asn Gly Leu Asp Leu Pro Pro Lys Gln	
360 365 370	
gtt tac atc gaa gtt tta att cta gat acc agc tta gag aaa tcc tgg	1267
Val Tyr Ile Glu Val Leu Ile Leu Asp Thr Ser Leu Glu Lys Ser Trp	
375 380 385	
gac ttt gga gtg caa tgg gta gcc cta ggt gat gaa caa agt aaa gta	1315
Asp Phe Gly Val Gln Trp Val Ala Leu Gly Asp Glu Gln Ser Lys Val	
390 395 400 405	

Fig. 15 con't)

gct tat gct tct gga cta ttg aat aat act ggc ata gcc aca cct aca	1363
Ala Tyr Ala Ser Gly Leu Leu Asn Asn Thr Gly Ile Ala Thr Pro Thr	
410 415 420	
aaa gca act gtc cct ccc ggc acg cca aat cct ggt tcg atc cct ctt	1411
Lys Ala Thr Val Pro Pro Gly Thr Pro Asn Pro Gly Ser Ile Pro Leu	
425 430 435	
cct acg cca gga caa ttg aca ggg ttc tca gat atg ctg aac tct tcg	1459
Pro Thr Pro Gly Gln Leu Thr Gly Phe Ser Asp Met Leu Asn Ser Ser	
440 445 450	
tca gca ttc ggt cta gga atc atc gga aat gtc cta agt cat aaa ggg	1507
Ser Ala Phe Gly Leu Gly Ile Ile Gly Asn Val Leu Ser His Lys Gly	
455 460 465	
aag tct ttc ctt act ttg gga ggc tta tta agt gcc tta gat caa gat	1555
Lys Ser Phe Leu Thr Leu Gly Gly Leu Leu Ser Ala Leu Asp Gln Asp	
470 475 480 485	
gga gat act gtc att gtc ttg aat cct aga atc atg gct cag gat acg	1603
Gly Asp Thr Val Ile Val Leu Asn Pro Arg Ile Met Ala Gln Asp Thr	
490 495 500	
caa caa gct tcg ttt ttt gta ggg caa acg gtc cct tac caa act atc	1651
Gln Gln Ala Ser Phe Phe Val Gly Gln Thr Val Pro Tyr Gln Thr Ile	
505 510 515	
aaa tac tat atc caa gaa aca gga act gta acg caa aat atc gat tat	1699
Lys Tyr Tyr Ile Gln Glu Thr Gly Thr Val Thr Gln Asn Ile Asp Tyr	
520 525 530	
gaa gat att gga gtg aac ctt gtc gtt acc tct aca gtt gct ccc aac	1747
Glu Asp Ile Gly Val Asn Leu Val Val Thr Ser Thr Val Ala Pro Asn	
535 540 545	
aat gta gtt aca cta caa atc gaa cag acg atc tca gaa tta cat tcc	1795
Asn Val Val Thr Leu Gln Ile Glu Gln Thr Ile Ser Glu Leu His Ser	
550 555 560 565	
gcg tct gga tca cta aca cct gtc aca gat aaa act tat gca gcc aca	1843
Ala Ser Gly Ser Leu Thr Pro Val Thr Asp Lys Thr Tyr Ala Ala Thr	
570 575 580	
cgc tta caa att ccc gac ggt tgt ttc tta gtt atg agt ggg cat atc	1891
Arg Leu Gln Ile Pro Asp Gly Cys Phe Leu Val Met Ser Gly His Ile	
585 590 595	
aga gat aaa act aca aaa gtg gtt tca gga gtg cct ttg cta aac tcc	1939
Arg Asp Lys Thr Thr Lys Val Val Ser Gly Val Pro Leu Leu Asn Ser	
600 605 610	

Fig. 15 con't)

ata cca tta att cgt ggt tta ttt agc cgt acc atc gac caa agg caa	1987
Ile Pro Leu Ile Arg Gly Leu Phe Ser Arg Thr Ile Asp Gln Arg Gln	
615 620 625	
aaa cgc aat atc atg atg ttt att aag cct aag gtg att agt agc ttt	2035
Lys Arg Asn Ile Met Met Phe Ile Lys Pro Lys Val Ile Ser Ser Phe	
630 635 640 645	
gaa gaa ggc act cgt gtt acc aat aag gaa gga tac aga tac aat tgg	2083
Glu Glu Gly Thr Arg Val Thr Asn Lys Glu Gly Tyr Arg Tyr Asn Trp	
650 655 660	
gaa gct gat gaa gga tcc atg caa gtg gcc cct cgc cat gct cct gaa	2131
Glu Ala Asp Glu Gly Ser Met Gln Val Ala Pro Arg His Ala Pro Glu	
665 670 675	
tgc caa gga cct cct tct tta cag gct gaa agt gac ttt aaa ata ata	2179
Cys Gln Gly Pro Pro Ser Leu Gln Ala Glu Ser Asp Phe Lys Ile Ile	
680 685 690	
gaa ata gaa gct cag tagtggtata taaaagagga agatgatatt ctccgccgtg	2234
Glu Ile Glu Ala Gln	
695	
gaatagcttc tgactctggt gcattcaggg ggaaagccaa gaagatgtag agtcggccgt	2294
ataact	2300

Figure 16 (RY-41)

## Restriction enzyme analysis of CPN100538

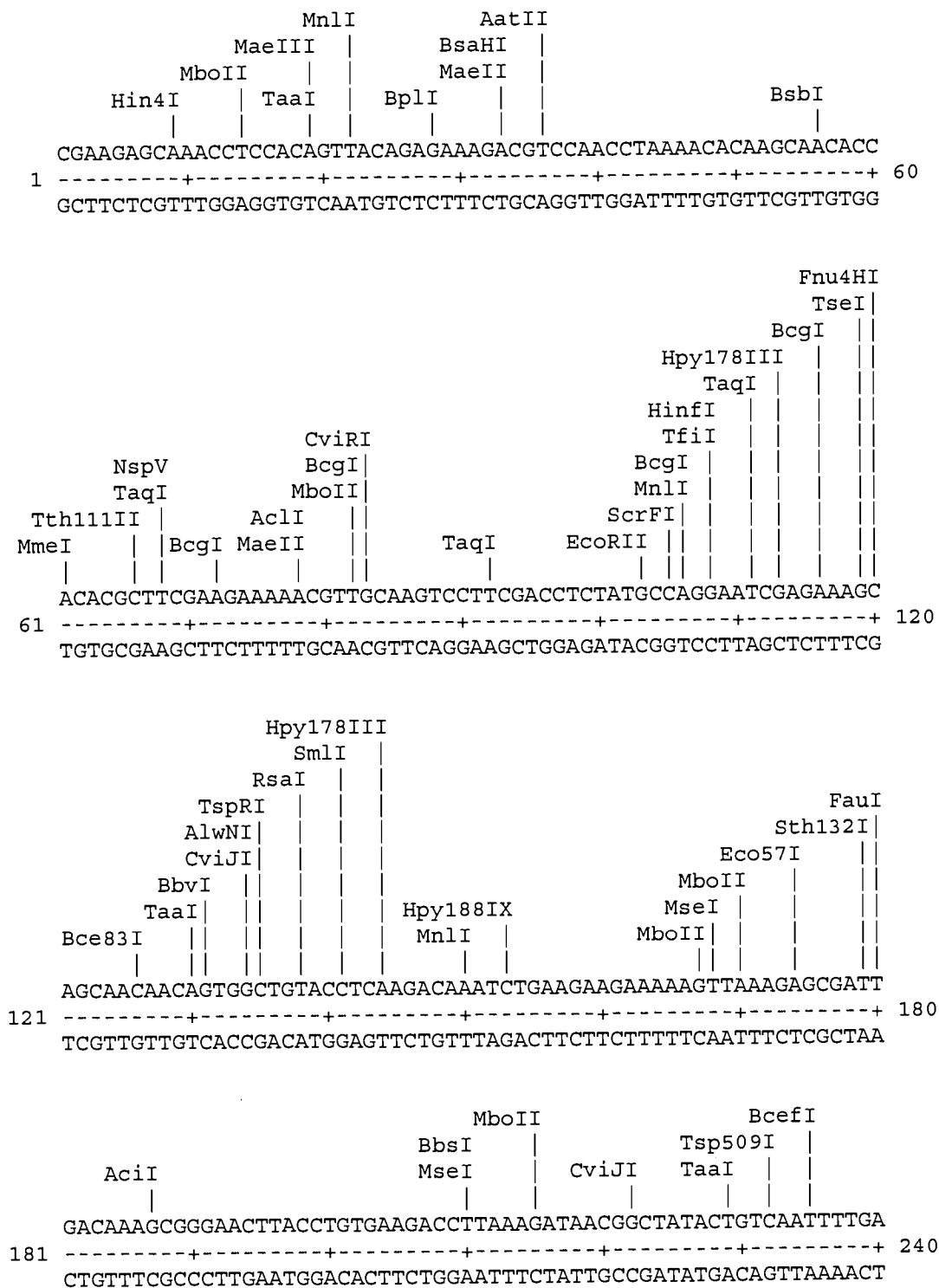
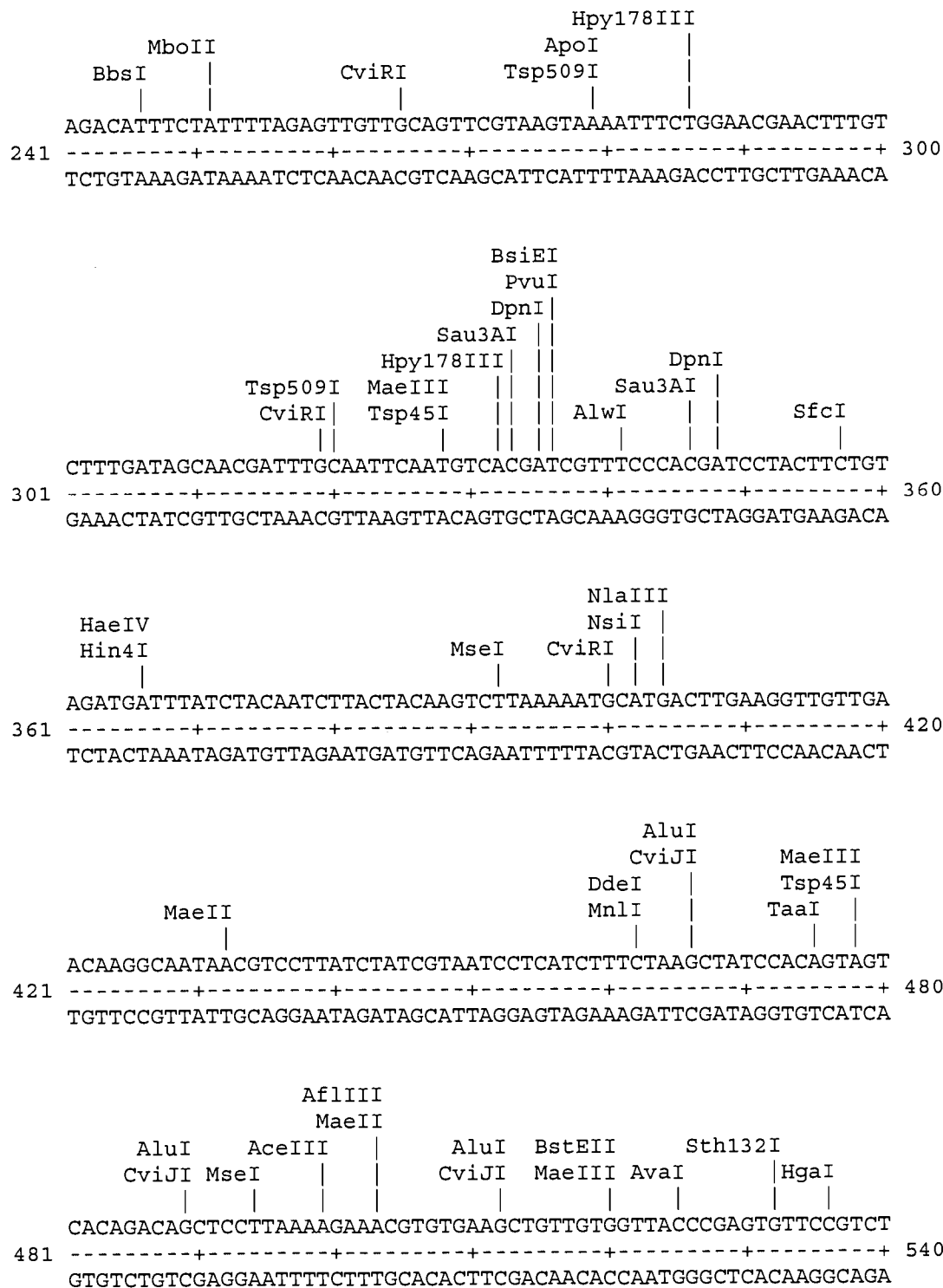


Fig. 16 (con't)



MnlI  
 PstI  
 Fnu4HI  
 CviRI  
 TseI  
 BbvI  
 SspI  
 SfaNI  
 Eco57I  
 NlaIII  
 BsaHI  
 CviJI  
 SfcI  
 541  
 TTACAGGCGTCAGCCCTCTGCAGCAGTAAATATTATTCAACCTTTACTTTCCCATGATGC  
 -----+-----+-----+-----+-----+-----+-----+-----+ 600  
 AATGTCCGCAGTCGGGAGACGTCGTCATTTATAATAAGTTGGAAATGAAAGGGTACTACG

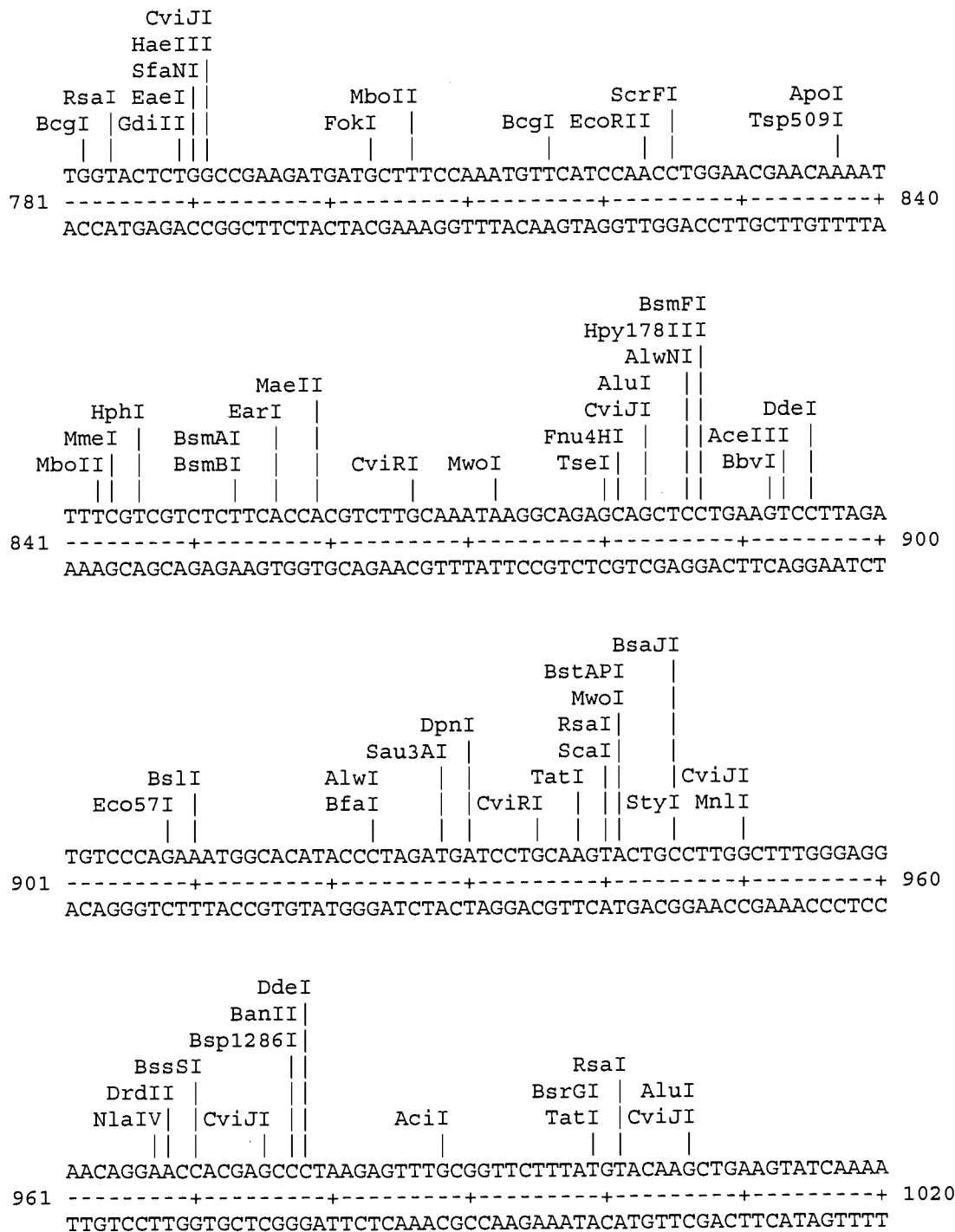
AluI  
 CviJI  
 MwoI  
 Hpy188IX  
 NlaIII  
 Hpy188IX  
 TaqI  
 601  
 TATCGTTAGTGCTTCAGAAGCTACTCGTCATGTTATCATCTCGGATATTGCTGGTAATGT  
 -----+-----+-----+-----+-----+-----+-----+-----+ 660  
 ATAGCAATCACGAAGTCTTCGATGAGCAGTACAATAGTAGAGCCTATAACGACCATTACA

Hpy178III  
 BfaI  
 XbaI  
 AluI  
 CviJI  
 Fnu4HI  
 TseI  
 Cac8I  
 BfaI  
 NheI  
 TspRI  
 ScrFI  
 BsaJI  
 EcoRII  
 AceIII  
 BbvI  
 MslI  
 NlaIII  
 661  
 CGATAAAGTCAGTGATTTGCTAGCAGCTCTAGATTGCCAGGCACATCTGTGGACATGAC  
 -----+-----+-----+-----+-----+-----+-----+-----+ 720  
 GCTATTTTCAGTCACTAAACGATCGTCGAGATCTAACGGGTCCGTGTAGACACCTGTACTG

FauI  
 Sth132I  
 AluI  
 CviJI  
 Fnu4HI  
 TseI  
 AciI  
 MwoI  
 MseI  
 AceIII  
 AluI  
 CviJI  
 BbvI  
 XcmI  
 721  
 TGAATACGAAGTTAAATATGCCAATCCCGCAGCTCTTGTTAGCTACTGCCAAGATGTTCT  
 -----+-----+-----+-----+-----+-----+-----+-----+ 780  
 ACTTATGCTTCAATTTATACGGTTAGGGCGTCGAGAACAATCGATGACGGTTCTACAAGA



Fig. 16 (con't)



Eco57I | MnlI | MaeIII | AluI  
 EcoRV | CviJI  
 MslI  
 TGGAGAAGTGATTGCTAATGCCCTCCAAGATATCGGTTACAATCTATATGTAACCACAGC  
 1021 -----+-----+-----+-----+-----+-----+ 1080  
 ACCTCTTCACTAACGATTACGGGAGGTTCTATAGCCAATGTTAGATATACATTGGTGTGC

BstXI | MboII | MnlI | BciVI  
 MseI | BsrI | TspRI  
 TATGGACGAAGATTTTCATTAACTCTCAATAGTATCCAGTGGTTAGAGGTCAATAACTC  
 1081 -----+-----+-----+-----+-----+-----+ 1140  
 ATACCTGCTTCTAAAGTAATTGTGAGAGTTATCATAGGTCACCAATCTCCAGTTATTGAG

HincII | CviJI  
 AccI | HaeI  
 BsaJI | StyI | SalI | HaeIII | MseI | MnlI  
 Hpy188IX | TaaI  
 CATAGTTATTATCGGAAACCAAGGGAATGTGCGACAGAGTTATTGGCCTCTTAAACGGTTT  
 1141 -----+-----+-----+-----+-----+-----+ 1200  
 GTATCAATAATAGCCTTTGGTTCCCTTACAGCTGTCTCAATAACCGGAGAATTTGCCAAA

Hpy178III | DdeI  
 Tsp509I | BfaI | AluI  
 CviJI  
 MnlI | TaqI | MseI | XbaI | CjePI  
 AGATTTACCTCCTAAACAGGTTTACATCGAAGTTTAAATTCTAGATACCAGCTTAGAGAA  
 1201 -----+-----+-----+-----+-----+-----+ 1260  
 TCTAAATGGAGGATTTGTCCAAATGTAGCTTCAAATTAAGATCTATGGTCTGAATCTCTT

BfaI  
 AvrII |  
 BsaJI |  
 ScrFI | BsmFI | StyI |  
 BsaJI | CjePI | CviJI |  
 EcoRII | CviRI | BsrDI | HphI | AluI  
 CviJI  
 ATCCTGGGACTTTGGAGTGCAATGGGTAGCCCTAGGTGATGAACAAAGTAAAGTAGCTTA  
 1261 -----+-----+-----+-----+-----+-----+ 1320  
 TAGGACCCTGAAACCTCACGTTACCCATCGGGATCCACTACTTGTTCATTTCATCGAAT

Fig. 16 (con't)

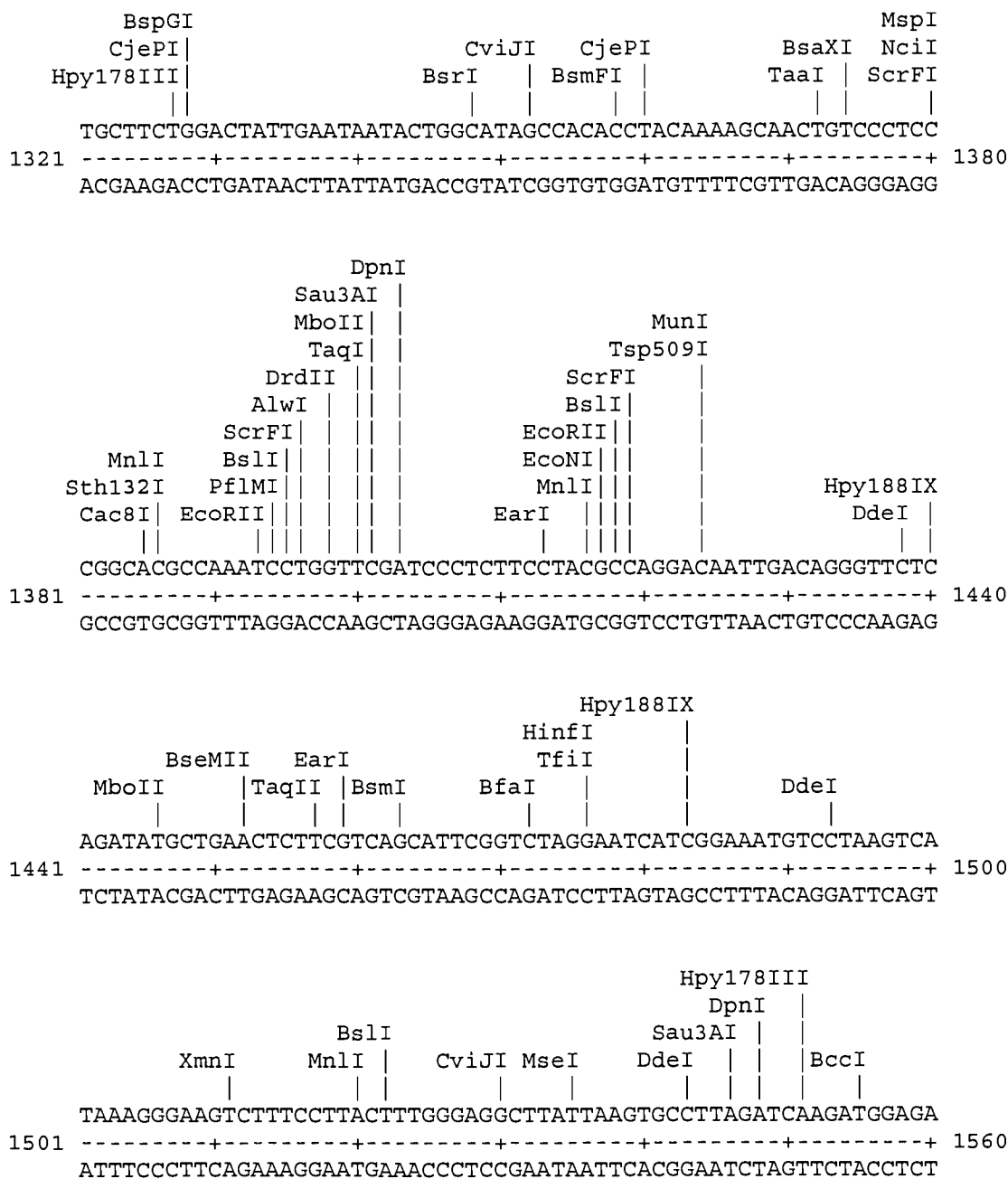


Fig. 16 (con't)

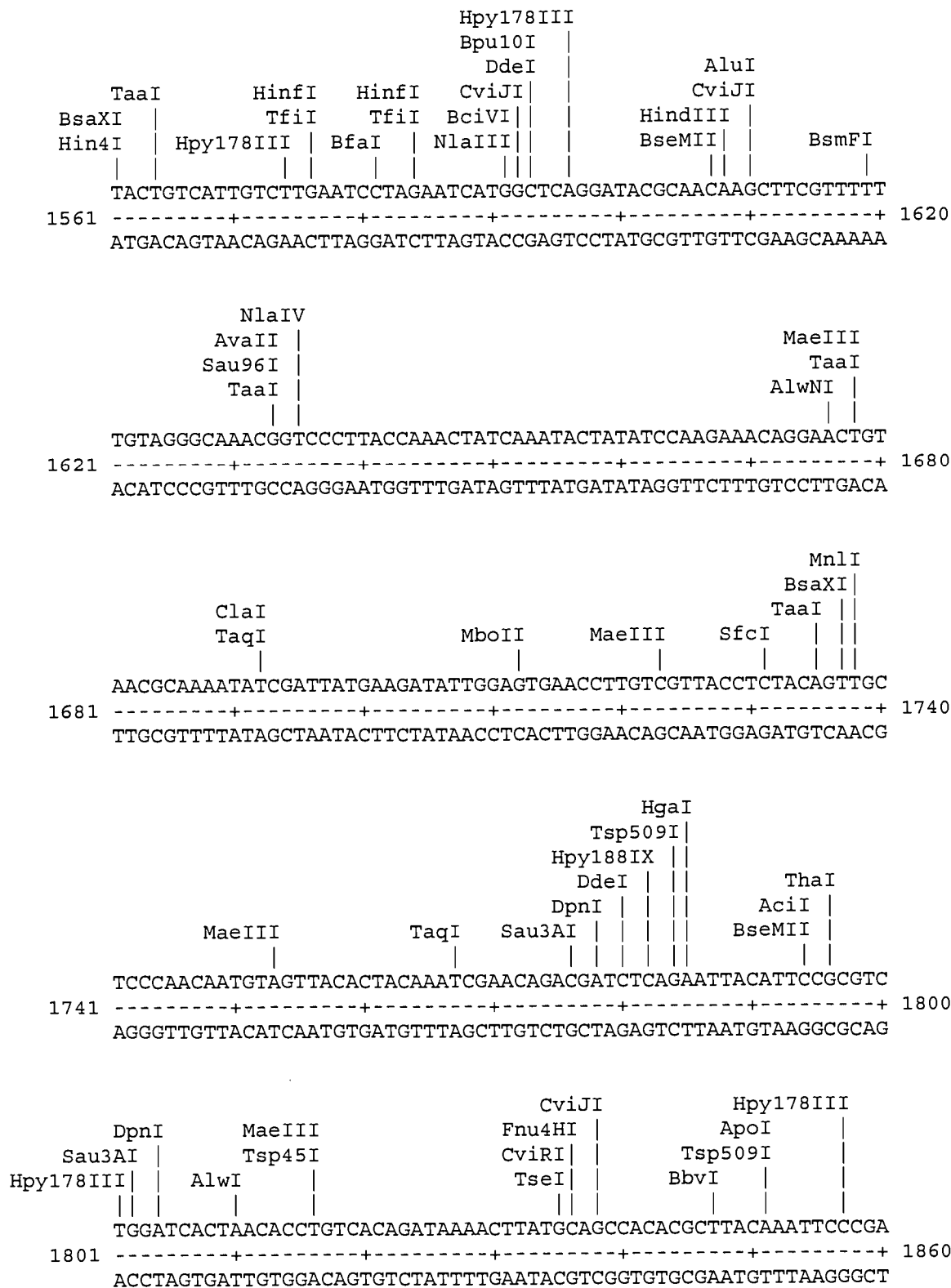


Fig. 16 (con't)

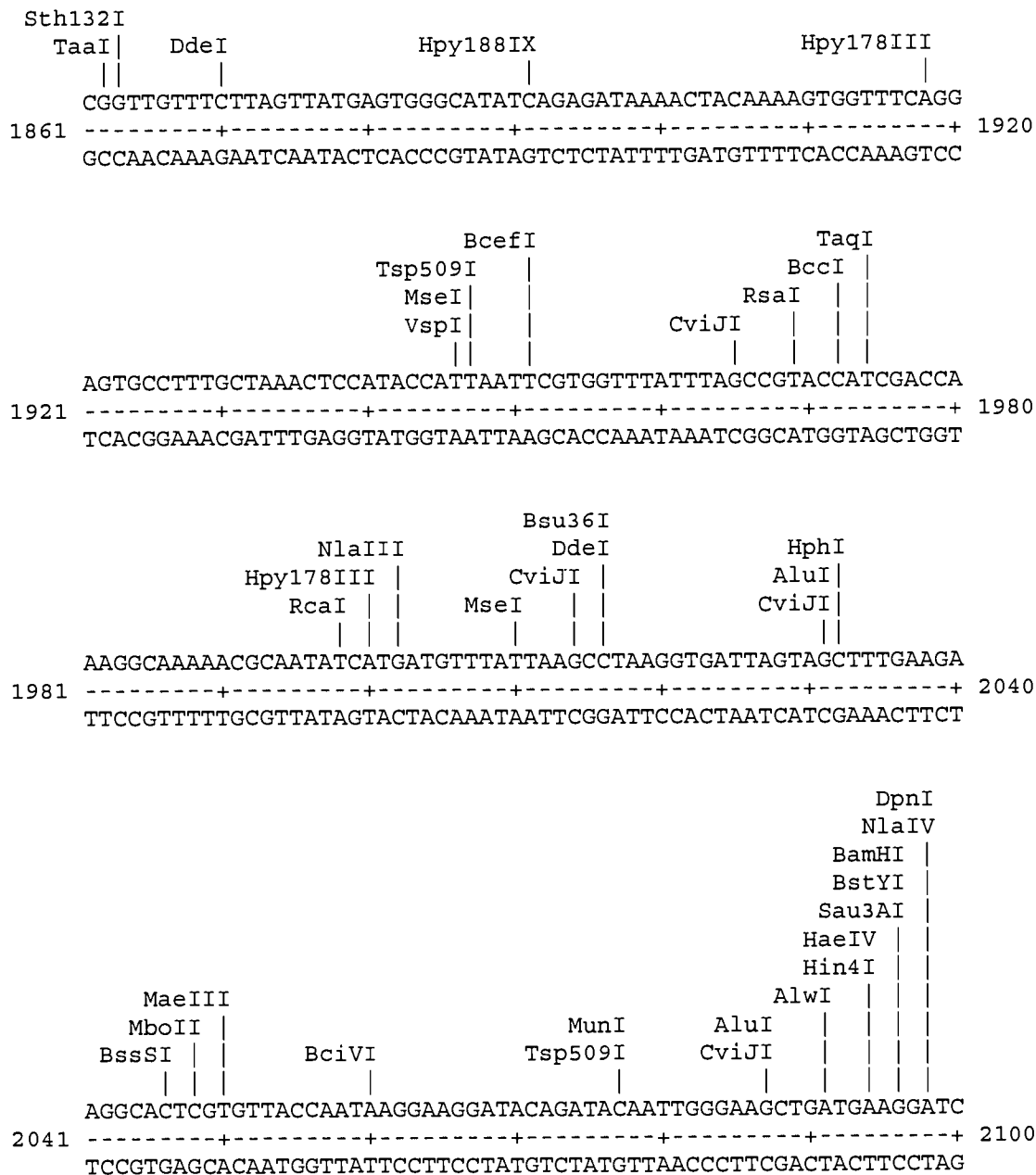


Fig. 16 (con't)

NlaIV  
 CviJI  
 HaeIII  
 Sau96I  
 BstXI  
 MslI  
 AlwI  
 CviRI  
 NlaIII  
 Hpy178III  
 MnlI  
 NlaIII  
 AvaII  
 EcoO109I  
 Psp5II  
 Sau96I  
 Sse8647I  
 BsmI  
 BsaJI  
 StyI  
 CviJI  
 MnlI  
 CATGCAAGTGGCCCTCGCCATGCTCCTGAATGCCAAGGACCTCCTTCTTTACAGGCTGA  
 2101 -----+-----+-----+-----+-----+-----+-----+ 2160  
 GTACGTTACCGGGAGCGGTACGAGGACTTACGGTTCTTGGAGGAAGAAATGTCCGACT

MaeIII  
 Tsp45I  
 DraI  
 MseI  
 DdeI  
 AluI  
 CviJI  
 BseMII  
 MnlI  
 Bcefi  
 AAGTGACTTTAAAATAATAGAAATAGAAGCTCAGTAGTGGTATATAAAAGAGGAAGATGA  
 2161 -----+-----+-----+-----+-----+-----+-----+ 2220  
 TTCACTGAAATTTTATTATCTTTATCTTCGAGTCATCACCATATATTTTCTCCTTCTACT

BsaJI  
 BstDSI  
 EciI  
 AciI  
 MboII  
 HinfI  
 Hpy188IX  
 AluI  
 CviJI  
 PleI  
 XmnI  
 BsmI  
 CviRI  
 CviJI  
 Bcefi  
 TATTCTCCGCCGTGGAATAGCTTCTGACTCTGTTGCATTCAGGGGAAAGCCAAGAAGAT  
 2221 -----+-----+-----+-----+-----+-----+-----+ 2280  
 ATAAGAGGCGGCACCTTATCGAAGACTGAGACAACGTAAGTCCCCCTTTTCGGTTCTTCTA

PleI  
 BsiEI  
 CviJI  
 HaeIII  
 EaeI  
 EagI  
 GdiII  
 MboII  
 HinfI  
 HaeIV  
 Hin4I  
 GTAGAGTCGGCCGTATAACT  
 2281 -----+-----+-----+-----+-----+-----+-----+ 2300  
 CATCTCAGCCGGCATATTGA

Figure 17: CPN100557

```

tagcttgaaa tagcttcctc caattgtgat ttctgaagaa gtataggggg aaatgtcgaa      60

gagatagtct tgttttaaag gaggagggga aaacggttta atg agc aga aaa gac      115
                               Met Ser Arg Lys Asp
                               Arg Lys Asp
                               1           5

aat gag gtt tcc tta gct cgt tca att ttt aat ata tta tcc gga act      163
Asn Glu Val Ser Leu Ala Arg Ser Ile Phe Asn Ile Leu Ser Gly Thr
Asn Glu Val Ser Leu Ala Arg Ser Ile Phe Asn Ile Leu Ser Gly Thr
                               10           15           20

ttc tgt agt cgt att aca ggg ata ttt cga gaa att gca atg gca acc      211
Phe Cys Ser Arg Ile Thr Gly Ile Phe Arg Glu Ile Ala Met Ala Thr
Phe Cys Ser Arg Ile Thr Gly Ile Phe Arg Glu Ile Ala Met Ala Thr
                               25           30           35

tat ttt gga gct gat cca att gta gct gct ttc tgg tta ggt ttc cgt      259
Tyr Phe Gly Ala Asp Pro Ile Val Ala Ala Phe Trp Leu Gly Phe Arg
Tyr Phe Gly Ala Asp Pro Ile Val Ala Ala Phe Trp Leu Gly Phe Arg
                               40           45           50

act gtt ttt ttc tta aga aaa att tta gga ggg ctc att cta gaa caa      307
Thr Val Phe Phe Leu Arg Lys Ile Leu Gly Gly Leu Ile Leu Glu Gln
Thr Val Phe Phe Leu Arg Lys Ile Leu Gly Gly Leu Ile Leu Glu Gln
                               55           60           65

gcc ttc atc cct cat ttt gaa ttt ctc cgt gct caa agt ctc gat cgt      355
Ala Phe Ile Pro His Phe Glu Phe Leu Arg Ala Gln Ser Leu Asp Arg
Ala Phe Ile Pro His Phe Glu Phe Leu Arg Ala Gln Ser Leu Asp Arg
                               70           75           80           85

gcg gcg ttt ttt ttc cga cgc ttt tct aga ttg att aaa ggc agc act      403
Ala Ala Phe Phe Phe Arg Arg Phe Ser Arg Leu Ile Lys Gly Ser Thr
Ala Ala Phe Phe Phe Arg Arg Phe Ser Arg Leu Ile Lys Gly Ser Thr
                               90           95           100

att ata ttc act ctg ctt att gaa gca gta ttg tgg gta ttc ttc aat      451
Ile Ile Phe Thr Leu Leu Ile Glu Ala Val Leu Trp Val Phe Phe Asn
Ile Ile Phe Thr Leu Leu Ile Glu Ala Val Leu Trp Val Phe Phe Asn
                               105           110           115

aac gtt gaa gag ggg act tac gat atg att ctc ctt act atg ata ctc      499
Asn Val Glu Glu Gly Thr Tyr Asp Met Ile Leu Leu Thr Met Ile Leu
Asn Val Glu Glu Gly Thr Tyr Asp Met Ile Leu Leu Thr Met Ile Leu
                               120           125           130

ttg ccc tgt ggc att ttc tta atg atg tac aat gta aac ggc gct ttg      547
Leu Pro Cys Gly Ile Phe Leu Met Met Tyr Asn Val Asn Gly Ala Leu
Leu Pro Cys Gly Ile Phe Leu Met Met Tyr Asn Val Asn Gly Ala Leu
                               135           140           145

ctt cac tgt gga aat aag ttt ttc ggg gtg gga tta gct ccc gta gtt      595
Leu His Cys Gly Asn Lys Phe Phe Gly Val Gly Leu Ala Pro Val Val
Leu His Cys Gly Asn Lys Phe Phe Gly Val Gly Leu Ala Pro Val Val
                               150           155           160           165

gta aat atc att tgg att ttc ttt gtt ata gcg gct cgt cat tca gat      643
Val Asn Ile Ile Trp Ile Phe Phe Val Ile Ala Ala Arg His Ser Asp
Val Asn Ile Ile Trp Ile Phe Phe Val Ile Ala Ala Arg His Ser Asp
                               170           175           180

```

Fig. 17 (con't)

cct aga gag cgt att atc ggt tta tcc gtg gct cta gtt atc ggg ttt	691
Pro Arg Glu Arg Ile Ile Gly Leu Ser Val Ala Leu Val Ile Gly Phe	
Pro Arg Glu Arg Ile Ile Gly Leu Ser Val Ala Leu Val Ile Gly Phe	
185 190 195	
ttc ttc gaa tgg tta atc acg gtt cct gga gta tgg aaa ttt cta tta	739
Phe Phe Glu Trp Leu Ile Thr Val Pro Gly Val Trp Lys Phe Leu Leu	
Phe Phe Glu Trp Leu Ile Thr Val Pro Gly Val Trp Lys Phe Leu Leu	
200 205 210	
gaa gcg aag agc cca cct caa gaa cac gat agt gtt cga gct tta tta	787
Glu Ala Lys Ser Pro Pro Gln Glu His Asp Ser Val Arg Ala Leu Leu	
Glu Ala Lys Ser Pro Pro Gln Glu His Asp Ser Val Arg Ala Leu Leu	
215 220 225	
gct ccc tta tct ttg ggt att tta act tca agc atc ttc cag ctg aac	835
Ala Pro Leu Ser Leu Gly Ile Leu Thr Ser Ser Ile Phe Gln Leu Asn	
Ala Pro Leu Ser Leu Gly Ile Leu Thr Ser Ser Ile Phe Gln Leu Asn	
230 235 240 245	
ctt ctt tct gat atc tgc ttg gct cgc tat gta cat gaa ata ggc cct	883
Leu Leu Ser Asp Ile Cys Leu Ala Arg Tyr Val His Glu Ile Gly Pro	
Leu Leu Ser Asp Ile Cys Leu Ala Arg Tyr Val His Glu Ile Gly Pro	
250 255 260	
cta tat ctt atg tac tcc tta aag att tat cag ctc ccc ata cat ctc	931
Leu Tyr Leu Met Tyr Ser Leu Lys Ile Tyr Gln Leu Pro Ile His Leu	
Leu Tyr Leu Met Tyr Ser Leu Lys Ile Tyr Gln Leu Pro Ile His Leu	
265 270 275	
ttt ggc ttt ggt gtg ttt acc gtt ctc ctc cca gca att tct cgt tgt	979
Phe Gly Phe Phe Gly Val Phe Thr Val Leu Leu Pro Ala Ile Ser Arg Cys	
Phe Gly Phe Phe Gly Val Phe Thr Val Leu Leu Pro Ala Ile Ser Arg Cys	
280 285 290	
gta cag cga gaa gat cat gag agg gga ttg aaa ctt atg aag ttc gtt	1027
Val Gln Arg Glu Asp His Glu Arg Gly Leu Lys Leu Met Lys Phe Val	
Val Gln Arg Glu Asp His Glu Arg Gly Leu Lys Leu Met Lys Phe Val	
295 300 305	
ctc acc cta acc atg tcc gta atg atc att atg aca gca ggg cta ttg	1075
Leu Thr Leu Thr Met Ser Val Met Ile Ile Met Thr Ala Gly Leu Leu	
Leu Thr Leu Thr Met Ser Val Met Ile Ile Met Thr Ala Gly Leu Leu	
310 315 320 325	
ctc tta gct tta cct gga gtc cgt gtc ctt tat gaa cac gga ctt ttc	1123
Leu Leu Ala Leu Pro Gly Val Arg Val Leu Tyr Glu His Gly Leu Phe	
Leu Leu Ala Leu Pro Gly Val Arg Val Leu Tyr Glu His Gly Leu Phe	
330 335 340	
cct cag agt gct gtc tac gct att gtt cgt gta ttg cga ggt tat ggt	1171
Pro Gln Ser Ala Val Tyr Ala Ile Val Arg Val Leu Arg Gly Tyr Gly	
Pro Gln Ser Ala Val Tyr Ala Ile Val Arg Val Leu Arg Gly Tyr Gly	
345 350 355	
gcc agt att atc cct atg gcc ttg gct cct tta gtc tct gtt ctt ttt	1219
Ala Ser Ile Ile Pro Met Ala Leu Ala Pro Leu Val Ser Val Leu Phe	
Ala Ser Ile Ile Pro Met Ala Leu Ala Pro Leu Val Ser Val Leu Phe	
360 365 370	



Fig. 17 (con't)

tat gca cag cgg cag tat gct gtt ccg ctc ttt ata gga atc ggt acg	1267
Tyr Ala Gln Arg Gln Tyr Ala Val Pro Leu Phe Ile Gly Ile Gly Thr	
Tyr Ala Gln Arg Gln Tyr Ala Val Pro Leu Phe Ile Gly Ile Gly Thr	
375 380 385	
gct ttg gcc aat att gtt tta agc ttg gtt cta ggt cgt tgg gtt tta	1315
Ala Leu Ala Asn Ile Val Leu Ser Leu Val Leu Gly Arg Trp Val Leu	
Ala Leu Ala Asn Ile Val Leu Ser Leu Val Leu Gly Arg Trp Val Leu	
390 395 400 405	
aaa gac gtc tcg ggc att tcc tat gct aca tcc ata act gct tgg gtg	1363
Lys Asp Val Ser Gly Ile Ser Tyr Ala Thr Ser Ile Thr Ala Trp Val	
Lys Asp Val Ser Gly Ile Ser Tyr Ala Thr Ser Ile Thr Ala Trp Val	
410 415 420	
cag tta tat ttc ctc tgg tat tat tct tcg aaa aga ctc cct atg tac	1411
Gln Leu Tyr Phe Leu Trp Tyr Tyr Ser Ser Lys Arg Leu Pro Met Tyr	
Gln Leu Tyr Phe Leu Trp Tyr Tyr Ser Ser Lys Arg Leu Pro Met Tyr	
425 430 435	
tct aag tta ctt tgg gag agc atc cgg cgt tcc ata aaa gtt atg gga	1459
Ser Lys Leu Leu Trp Glu Ser Ile Arg Arg Ser Ile Lys Val Met Gly	
Ser Lys Leu Leu Trp Glu Ser Ile Arg Arg Ser Ile Lys Val Met Gly	
440 445 450	
acc act atg ctt gct tgt atg att act cta ggc tta aat atc ctt acg	1507
Thr Thr Met Leu Ala Cys Met Ile Thr Leu Gly Leu Asn Ile Leu Thr	
Thr Thr Met Leu Ala Cys Met Ile Thr Leu Gly Leu Asn Ile Leu Thr	
455 460 465	
caa act aca tat gta att ttc tta aac ccc ctc aca cca ctt gct tgg	1555
Gln Thr Thr Tyr Val Ile Phe Leu Asn Pro Leu Thr Pro Leu Ala Trp	
Gln Thr Thr Tyr Val Ile Phe Leu Asn Pro Leu Thr Pro Leu Ala Trp	
470 475 480 485	
ccc tta tcc tcc ata acg gct caa gca att gct ttt tta tct gag agc	1603
Pro Leu Ser Ser Ile Thr Ala Gln Ala Ile Ala Phe Leu Ser Glu Ser	
Pro Leu Ser Ser Ile Thr Ala Gln Ala Ile Ala Phe Leu Ser Glu Ser	
490 495 500	
tgc att ttc ttg gct ttt ttg ttt ggt ttt gca aaa ctg ctt cga gta	1651
Cys Ile Phe Leu Ala Phe Leu Phe Gly Phe Ala Lys Leu Leu Arg Val	
Cys Ile Phe Leu Ala Phe Leu Phe Gly Phe Ala Lys Leu Leu Arg Val	
505 510 515	
gaa gat ctt att aat ttg gct tct ttt gaa tac tgg cgt ggg caa cgg	1699
Glu Asp Leu Ile Asn Leu Ala Ser Phe Glu Tyr Trp Arg Gly Gln Arg	
Glu Asp Leu Ile Asn Leu Ala Ser Phe Glu Tyr Trp Arg Gly Gln Arg	
520 525 530	
ggt ctt ttg caa aga caa cac gtg atg caa gac act caa aat	1741
Gly Leu Leu Gln Arg Gln His Val Met Gln Asp Thr Gln Asn	
Gly Leu Leu Gln	
535 540 545	
taatcatggt tgtttcttgt agctcagtcg ctttctttta gctttaagtt ttgatagcct	1801
gcttgggtctt ctgtttctac acttaatat gatactaagg atactatgaa aaaacaggta	1861
tatcaatggt tagcgagtgt gggtctttta gcgctgaca	1900

Figure 18 (RY-43)

## Restriction enzyme analysis of CPN100557

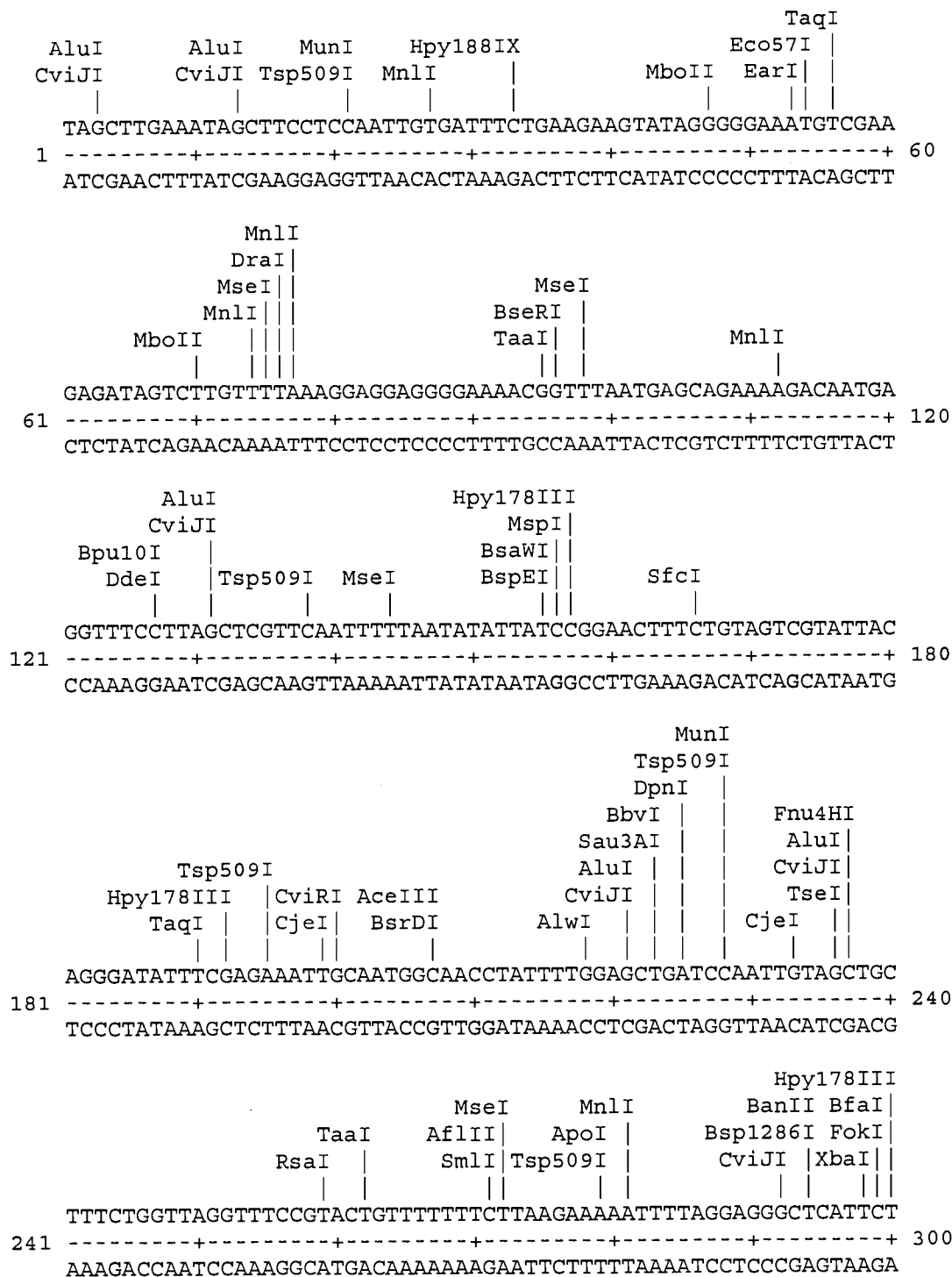
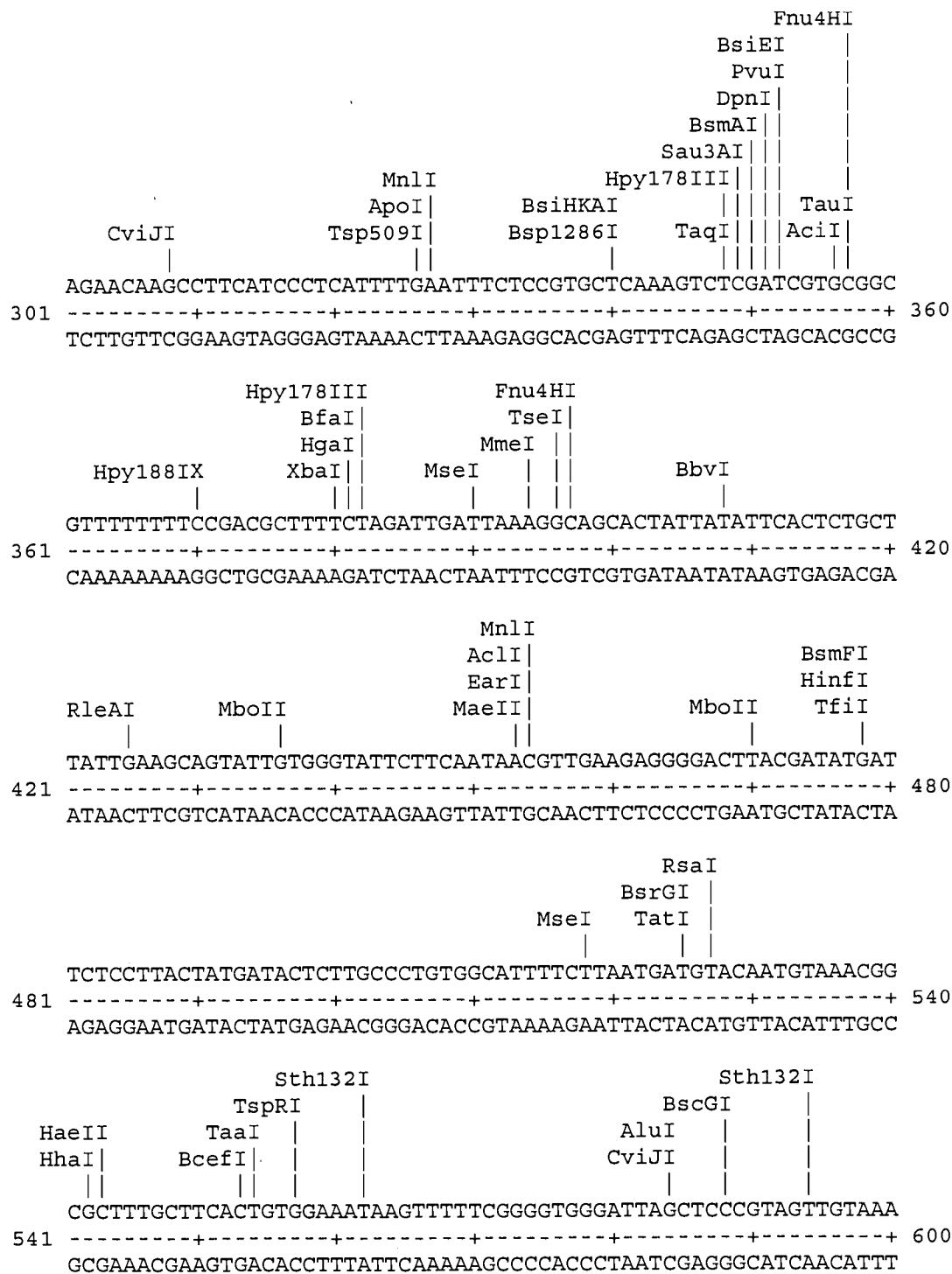


Fig. 18 (con't)



TATCATTGATTTCTTTGTTATAGCGGCTCGTCATTCAGATCCTAGAGAGCGTATTAT  
 601 -----+-----+-----+-----+-----+-----+-----+ 660  
 ATAGTAAACCTAAAGAAACAATATCGCCGAGCAGTAAGTCTAGGATCTCTCGCATAATA

BfaI  
 CviJI BstYI  
 Fnu4HI Sau3AI  
 TauI Hpy188IX  
 AciI AlwI

BfaI  
 Sth132I  
 CviJI  
 BsaJI  
 BstDSI  
 MboII  
 NspV  
 TaqI  
 MseI  
 ScrFI  
 EcoRII  
 NlaIV  
 TaaI

CGGTTTATCCGTGGCTCTAGTTATCGGGTTTTTCTTCGAATGGTTAATCACGGTTCCTGG  
 661 -----+-----+-----+-----+-----+-----+-----+ 720  
 GCCAAATAGGCACCGAGATCAATAGCCCAAAAAGAAGCTTACCAATTAGTGCCAAGGACC

Hpy178III  
 MboII  
 Bce83I  
 EarI  
 BanII  
 SapI  
 Bsp1286I  
 ApoI  
 BpmI  
 Hin4I  
 Tsp509I  
 CviJI  
 SmlI  
 MnlI  
 BplI  
 MslI  
 CviJI  
 AluI  
 TaqI

AGTATGGAAATTTCTATTAGAAGCGAAGAGCCACCTCAAGAACACGATAGTGTTCGAGC  
 721 -----+-----+-----+-----+-----+-----+-----+ 780  
 TCATACCTTTAAAGATAATCTTCGCTTCTCGGGTGGAGTCTTGTGCTATCACAAGCTCG

Tth111III  
 AluI  
 CviJI  
 MspAII  
 PvuII  
 SfaNI  
 MboII  
 MseI

AluI  
 CviJI  
 MwoI

TTTATTAGCTCCCTTATCTTTGGGTATTTTAACTTCAAGCATCTTCCAGCTGAACCTTCT  
 781 -----+-----+-----+-----+-----+-----+-----+ 840  
 AAATAATCGAGGGAATAGAAACCCATAAAATTGAAGTTCGTAGAAGGTCGACTTGGAAGA

CviJI  
 HaeIII  
 EcoO109I  
 NlaIII  
 RsaI  
 EcoRV  
 Cac8I  
 MwoI  
 BsrGI  
 TatI  
 Sau96I  
 MnlI  
 RsaI  
 TatI

Hpy188IX  
 Tth111III

TTCTGATATCTGCTTGGCTCGCTATGTACATGAAATAGGCCCTCTATATCTTATGTACTC  
 841 -----+-----+-----+-----+-----+-----+-----+ 900  
 AAGACTATAGACGAACCGAGCGATACATGTACTTTATCCGGGAGATATAGAATACATGAG

MseI                      AluI                      AceIII                      CviJI                      BseRI TaaI

CTTAAAGATTTATCAGCTCCCCATACATCTCTTTGGCTTTGGTGTGTTTTACCGTTCTCCT  
901 -----+-----+-----+-----+-----+-----+-----+ 960  
GAATTTCTAAATAGTCGAGGGGTATGTAGAGAAACCGAAACCACACAAATGGCAAGAGGA

MboII

NlaIII

Hpy178III

RcaI

DpnI

MnlI

RsaI

BsrGI

TatI

Sau3AI

MnlI

Tsp509I

CCCAGCAATTTCTCGTTGTGTACAGCGAGAAGATCATGAGAGGGGATTGAAACTTATGAA  
961 -----+-----+-----+-----+-----+-----+-----+ 1020  
GGGTCGTTAAAGAGCAACACATGTCGCTCTTCTAGTACTCTCCCTAACTTTGAATACTT

DpnI

BclI

Sau3AI

NlaIII

HphI

CviJI

DdeI

GTCGTTCTCACCCCTAACCATGTCCGTAATGATCATTATGACAGCAGGGCTATTGCTCTT  
1021 -----+-----+-----+-----+-----+-----+-----+ 1080  
CAAGCAAGAGTGGGATTGGTACAGGCATTACTAGTAATACTGTCTGCCGATAACGAGAA

BsaXI

Hin4I

HinfI

ScrFI

EcoRII

AluI

CviJI

PleI

BpmI

Hpy188IX

DdeI

MnlI

BseMII

MwoI

AccI

AGCTTTACCTGGAGTCCGTGTCTTTTATGAACACGGACTTTTCCCTCAGAGTGCTGTCTA  
1081 -----+-----+-----+-----+-----+-----+-----+ 1140  
TCGAAATGGACCTCAGGCACAGGAAATACTTGTGCCTGAAAAGGGAGTCTCACGACAGAT

BsaJI

StyI

BsrI

NlaIV

BanI

MnlI

CviJI

HaeI

NlaIV

HaeIII

CviJI

CGCTATTGTTCGTGTATTGCGAGGTTATGGTGCCAGTATTATCCCTATGGCCTTGGCTCC  
1141 -----+-----+-----+-----+-----+-----+-----+ 1200  
GCGATAACAAGCACATAACGCTCCAATACCACGGTCATAATAGGGATACCGGAACCGAGG

Fnu4HI  
 TauI  
 AciI  
 MspA1I  
 BsmAI  
 CviRI  
 BsrBI  
 AciI  
 HinfI  
 TfiI  
 1201  
 TTTAGTCTCTGTTCTTTTTTATGCACAGCGGCAGTATGCTGTTCCGCTCTTTATAGGAAT  
 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 1260  
 AAATCAGAGACAAGAAAAAATACGTGTCGCCGTATACGACAAGGCGAGAAATATCCTTA  
 CviJI  
 HaeI  
 HaeIII  
 CviJI  
 MscI  
 BceI  
 HindIII  
 CviJI  
 BsaHI  
 MaeII  
 RsaI  
 EaeI  
 SspI  
 MseI  
 DrdII  
 BfaI  
 Sth132I  
 DraI  
 MseI  
 1261  
 CGGTACGGCTTTGGCCAATATTGTTTTAAGCTTGGTCTAGGTCGTTGGGTTTTAAAAGA  
 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 1320  
 GCCATGCCGAAACCGGTTATAACAAAATTGAAACCAAGATCCAGCAACCCAAAATTTTCT  
 FokI  
 BsmAI  
 BsmBI  
 AvaI  
 AatII  
 Tth111III  
 TaqII  
 CviRI  
 MboII  
 1321  
 CGTCTCGGGCATTTCCTATGCTACATCCATAACTGCTTGGGTGCAGTTATATTTTCCTCTG  
 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 1380  
 GCAGAGCCCGTAAAGGATACGATGTAGGTATTGACGAACCCACGTCAATATAAAGGAGAC  
 PleI  
 MnlI  
 NspV  
 BsgI  
 TaqI  
 HinfI  
 RsaI  
 MaeIII  
 FokI  
 TatI  
 DdeI  
 SfaNI  
 MspI  
 1381  
 GTATTATTCTTCGAAAAGACTCCCTATGTACTCTAAGTTACTTTGGGAGAGCATCCGGCG  
 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 1440  
 CATAATAAGAAGCTTTTCTGAGGGATACATGAGATTCAATGAAACCCCTCTCGTAGGCCCGC  
 Tth111III  
 DrdII  
 NlaIV  
 Tth111III  
 XcmI  
 Cac8I  
 MseI  
 CviJI  
 BfaI  
 1441  
 TTCCATAAAAGTTATGGGAACCACTATGCTTGCTTGTATGATTACTCTAGGCTTAAATAT  
 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 1500  
 AAGGTATTTTCAATACCCTTGGTGATACGAACGAACATACTAATGAGATCCGAATTTATA

Fig. 18 (con't)

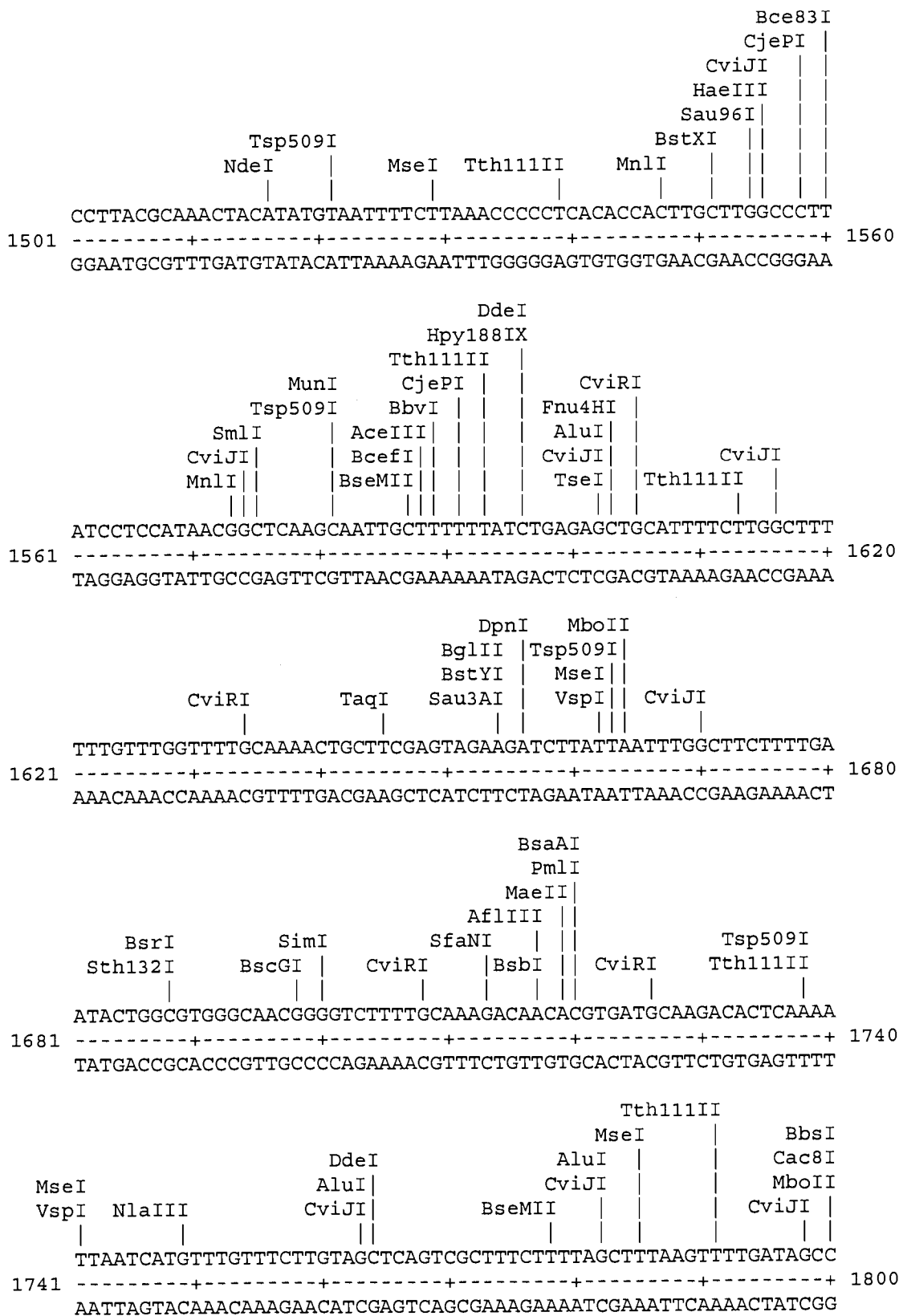


Fig. 18 (con't)

```

                SspI   DdeI
                |     |
            MseI | BciVI |
                |     |
1801  TGCTTGGTCTTCTGTTTCTACACTTAATATTGATACTAAGGATACTATGAAAAACAGGT
-----+-----+-----+-----+-----+-----+-----+ 1860
      ACGAACCAGAAGACAAAGATGTGAATTATAACTATGATTCCTATGATACTTTTTTGTCCA

                                HaeII
                                |
                                HhaI
                                |
                        DrdII  Eco47III
                        |     |||
1861  ATATCAATGGTTAGCGAGTGTGGTTCTTTTAGCGCTGACA
-----+-----+-----+-----+ 1900
      TATAGTTACCAATCGCTCACACCAAGAAAATCGCGACTGT
```



Figure 19: CPN100622

```

tctcaagagt aaccttatcc ttagattatt cagctcaagt ctcctcgtca actgtaggtc 60
aataccttaa agctgagagt cattgcacat tttaaccaca atg aaa aca tca agg 115
                                         Met Lys Thr Ser Arg
                                         1           5
aat aaa cag tgc aaa ata aca gat ccc tta agt aaa tct tcc ttc ttt 163
Asn Lys Gln Cys Lys Ile Thr Asp Pro Leu Ser Lys Ser Ser Phe Phe
              10           15           20
gtt gga gcc tta att tta ggt aaa act aca ata ctc ctt aat gcg act 211
Val Gly Ala Leu Ile Leu Gly Lys Thr Thr Ile Leu Leu Asn Ala Thr
              25           30           35
ccg ttg tct gac tat ttt gat aat caa gca aat caa ctc aca aca ctc 259
Pro Leu Ser Asp Tyr Phe Asp Asn Gln Ala Asn Gln Leu Thr Thr Leu
              40           45           50
ttc cct cta att gat act ctt act aac atg act ccc tac tct cat aga 307
Phe Pro Leu Ile Asp Thr Leu Thr Asn Met Thr Pro Tyr Ser His Arg
              55           60           65
gca aca ctt ttt gga gtt agg gat gac act aac caa gac att gtc ctc 355
Ala Thr Leu Phe Gly Val Arg Asp Asp Thr Asn Gln Asp Ile Val Leu
              70           75           80           85
gat cac cag aat tcc ata gaa agc tgg ttc gaa aac ttc tct caa gac 403
Asp His Gln Asn Ser Ile Glu Ser Trp Phe Glu Asn Phe Ser Gln Asp
              90           95           100
ggc ggt gct ctc tct tgc aaa tca ctt gcc ata acg aat aca aaa aac 451
Gly Gly Ala Leu Ser Cys Lys Ser Leu Ala Ile Thr Asn Thr Lys Asn
              105           110           115
caa att ctt ttc cta aat agc ttt gct att aaa aga gct ggt gcg atg 499
Gln Ile Leu Phe Leu Asn Ser Phe Ala Ile Lys Arg Ala Gly Ala Met
              120           125           130
tat gtt gat ggt aat ttc gat ctt tct gag aat cat ggt tcc atc att 547
Tyr Val Asp Gly Asn Phe Asp Leu Ser Glu Asn His Gly Ser Ile Ile
              135           140           145
ttc tct ggg aat tta agc ttt cct aat gca agt aat ttc gct gat act 595
Phe Ser Gly Asn Leu Ser Phe Pro Asn Ala Ser Asn Phe Ala Asp Thr
              150           155           160           165
tgt aca ggg gga gct gtt tta tgt tgc aaa aat gtt aca atc tca aaa 643
Cys Thr Gly Gly Ala Val Leu Cys Ser Lys Asn Val Thr Ile Ser Lys
      Thr Gly Gly Ala Val Leu Cys Ser Lys Asn Val Thr Ile Ser Lys
              170           175           180
aat caa gga acc gca tac ttc att aac aac aag gca aaa tct tca gga 691
Asn Gln Gly Thr Ala Tyr Phe Ile Asn Asn Lys Ala Lys Ser Ser Gly
Asn Gln Gly Thr Ala Tyr Phe Ile Asn Asn Lys Ala Lys Ser Ser Gly
              185           190           195

```

Fig. 19 (con't)

gga gca atc caa gct gca atc ata aac att aag gac aac act ggc cct	739
Gly Ala Ile Gln Ala Ala Ile Ile Asn Ile Lys Asp Asn Thr Gly Pro	
Gly Ala Ile Gln Ala Ala Ile Ile Asn Ile Lys Asp Asn Thr Gly Pro	
200 205 210	
tgc ctg ttt ttt aat aat gct gca ggc gga aca gcg ggg ggc gcg ttg	787
Cys Leu Phe Phe Asn Asn Ala Ala Gly Gly Thr Ala Gly Gly Ala Leu	
Cys Leu Phe Phe Asn Asn Ala Ala Gly Gly Thr Ala Gly Gly Ala Leu	
215 220 225	
ttc gct aat gct tgt aga att gag aat aat tct cag cct atc tat ttt	835
Phe Ala Asn Ala Cys Arg Ile Glu Asn Asn Ser Gln Pro Ile Tyr Phe	
Phe Ala Asn Ala Cys Arg Ile Glu Asn Asn Ser Gln Pro Ile Tyr Phe	
230 235 240 245	
ttg aat aac caa tca ggt ctg ggt ggt gca ata aga gta cat caa gag	883
Leu Asn Asn Gln Ser Gly Leu Gly Gly Ala Ile Arg Val His Gln Glu	
Leu Asn Asn Gln Ser Gly Leu Gly Gly Ala Ile Arg Val His Gln Glu	
250 255 260	
tgc att ctt aca aag aat acc ggt tct gtg atc ttc aac aat aat ttt	931
Cys Ile Leu Thr Lys Asn Thr Gly Ser Val Ile Phe Asn Asn Asn Phe	
Cys Ile Leu Thr Lys Asn Thr Gly Ser Val Ile Phe Asn Asn Asn Phe	
265 270 275	
gcc atg gaa gcg gac atc tct gct aac cat tcc tct gga ggg gct atc	979
Ala Met Glu Ala Asp Ile Ser Ala Asn His Ser Ser Gly Gly Ala Ile	
Ala Met Glu Ala Asp Ile Ser Ala Asn His Ser Ser Gly Gly Ala Ile	
280 285 290	
tat tgc att agt tgt tct ata aaa gac aac cca gga att gca gcc ttc	1027
Tyr Cys Ile Ser Cys Ser Ile Lys Asp Asn Pro Gly Ile Ala Ala Phe	
Tyr Cys Ile Ser Cys Ser Ile Lys Asp Asn Pro Gly Ile Ala Ala Phe	
295 300 305	
gat aat aat act gca gca cga gat gga ggt gct atc tgt aca caa tct	1075
Asp Asn Asn Thr Ala Ala Arg Asp Gly Gly Ala Ile Cys Thr Gln Ser	
Asp Asn Asn Thr Ala Ala Arg Asp Gly Gly Ala Ile Cys Thr Gln Ser	
310 315 320 325	
cta act ata caa gac agt ggt ccc gtc tat ttc aca aac aat cag gga	1123
Leu Thr Ile Gln Asp Ser Gly Pro Val Tyr Phe Thr Asn Asn Gln Gly	
Leu Thr Ile Gln Asp Ser Gly Pro Val Tyr Phe Thr Asn Asn Gln Gly	
330 335 340	
act tgg ggc ggc gct atc atg ctc cgt caa gat ggt gca tgc act tta	1171
Thr Trp Gly Gly Ala Ile Met Leu Arg Gln Asp Gly Ala Cys Thr Leu	
Thr Trp Gly Gly Ala Ile Met Leu Arg Gln Asp Gly Ala Cys Thr Leu	
345 350 355	
ttt gct gat cag gga gat att att ttt tat aat aat aga cac ttc aaa	1219
Phe Ala Asp Gln Gly Asp Ile Ile Phe Tyr Asn Asn Arg His Phe Lys	
Phe Ala Asp Gln Gly Asp Ile Ile Phe Tyr Asn Asn Arg His Phe Lys	
360 365 370	
gat act ttc agc aat cat gtt tct gta aac tgc acg cgt aat gtc tca	1267
Asp Thr Phe Ser Asn His Val Ser Val Asn Cys Thr Arg Asn Val Ser	
Asp Thr Phe Ser Asn His Val Ser Val Asn Cys Thr Arg Asn Val Ser	
375 380 385	

Fig. 19 (con't)

tta	aca	gtt	gga	gca	agt	caa	ggt	cat	tct	gct	acc	ttc	tat	gat	ccc	1315
Leu	Thr	Val	Gly	Ala	Ser	Gln	Gly	His	Ser	Ala	Thr	Phe	Tyr	Asp	Pro	
Leu	Thr	Val	Gly	Ala	Ser	Gln	Gly	His	Ser	Ala	Thr	Phe	Tyr	Asp	Pro	
390					395				400						405	
ata	cta	caa	aga	tat	act	ata	caa	aac	tct	atc	caa	aaa	ttt	aat	cct	1363
Ile	Leu	Gln	Arg	Tyr	Thr	Ile	Gln	Asn	Ser	Ile	Gln	Lys	Phe	Asn	Pro	
Ile	Leu	Gln	Arg	Tyr	Thr	Ile	Gln	Asn	Ser	Ile	Gln	Lys	Phe	Asn	Pro	
			410						415					420		
aat	cca	gaa	cac	ctc	gga	act	atc	ttg	ttc	tcc	tca	aca	tat	att	cgg	1411
Asn	Pro	Glu	His	Leu	Gly	Thr	Ile	Leu	Phe	Ser	Ser	Thr	Tyr	Ile	Pro	
Asn	Pro	Glu	His	Leu	Gly	Thr	Ile	Leu	Phe	Ser	Ser	Thr	Tyr	Ile	Pro	
			425					430					435			
gat	aca	tcg	act	tct	cgt	gat	gac	ttc	att	tca	cat	ttc	aga	aac	cac	1459
Asp	Thr	Ser	Thr	Ser	Arg	Asp	Asp	Phe	Ile	Ser	His	Phe	Arg	Asn	His	
Asp	Thr	Ser	Thr	Ser	Arg	Asp	Asp	Phe	Ile	Ser	His	Phe	Arg	Asn	His	
			440					445				450				
att	gga	ctg	tac	aac	ggc	aca	ctc	gct	ctt	gaa	gat	cga	gca	gag	tgg	1507
Ile	Gly	Leu	Tyr	Asn	Gly	Thr	Leu	Ala	Leu	Glu	Asp	Arg	Ala	Glu	Trp	
Ile	Gly	Leu	Tyr	Asn	Gly	Thr	Leu	Ala	Leu	Glu	Asp	Arg	Ala	Glu	Trp	
	455					460						465				
aaa	gtc	tat	aaa	ttt	gat	caa	ttt	ggt	ggg	act	cta	cgg	tta	ggc	agt	1555
Lys	Val	Tyr	Lys	Phe	Asp	Gln	Phe	Gly	Gly	Thr	Leu	Arg	Leu	Gly	Ser	
Lys	Val	Tyr	Lys	Phe	Asp	Gln	Phe	Gly	Gly	Thr	Leu	Arg	Leu	Gly	Ser	
470					475					480					485	
aga	gct	gtg	ttt	tct	aca	aca	gac	gaa	gaa	caa	agt	agc	agt	agt	gtg	1603
Arg	Ala	Val	Phe	Ser	Thr	Thr	Asp	Glu	Glu	Gln	Ser	Ser	Ser	Ser	Val	
Arg	Ala	Val	Phe	Ser	Thr	Thr	Asp	Glu	Glu	Gln	Ser	Ser	Ser	Ser	Val	
			490							495					500	
ggt	tct	gta	att	aac	atc	aat	aat	ctt	gca	att	aac	ctt	ccc	tct	atc	1651
Gly	Ser	Val	Ile	Asn	Ile	Asn	Asn	Leu	Ala	Ile	Asn	Leu	Pro	Ser	Ile	
Gly	Ser	Val	Ile	Asn	Ile	Asn	Asn	Leu	Ala	Ile	Asn	Leu	Pro	Ser	Ile	
			505					510					515			
tta	ggc	aac	aga	gtt	gct	ccc	aag	cta	tgg	att	cgc	ccc	aca	ggt	tca	1699
Leu	Gly	Asn	Arg	Val	Ala	Pro	Lys	Leu	Trp	Ile	Arg	Pro	Thr	Gly	Ser	
Leu	Gly	Asn	Arg	Val	Ala	Pro	Lys	Leu	Trp	Ile	Arg	Pro	Thr	Gly	Ser	
			520				525					530				
tca	gca	ccc	tat	agc	gaa	gat	aat	aac	cct	ata	atc	aat	ctc	tca	gga	1747
Ser	Ala	Pro	Tyr	Ser	Glu	Asp	Asn	Asn	Pro	Ile	Ile	Asn	Leu	Ser	Gly	
Ser	Ala	Pro	Tyr	Ser	Glu	Asp	Asn	Asn	Pro	Ile	Ile	Asn	Leu	Ser	Gly	
	535					540					545					
cct	ttg	agc	cta	ctg	gat	gac	gag	aac	cta	gat	ccc	tat	gat	act	gca	1795
Pro	Leu	Ser	Leu	Leu	Asp	Asp	Glu	Asn	Leu	Asp	Pro	Tyr	Asp	Thr	Ala	
Pro	Leu	Ser	Leu	Leu	Asp	Asp	Glu	Asn	Leu	Asp	Pro	Tyr	Asp	Thr	Ala	
550					555					560					565	
gac	ctt	gcc	caa	cct	atc	gca	gaa	gtt	cct	ctt	ctg	tat	ctc	tta	gac	1843
Asp	Leu	Ala	Gln	Pro	Ile	Ala	Glu	Val	Pro	Leu	Leu	Tyr	Leu	Leu	Asp	
Asp	Leu	Ala	Gln	Pro	Ile	Ala	Glu	Val	Pro	Leu	Leu	Tyr	Leu	Leu	Asp	
			570						575						580	

Fig. 19 (con't)

tac agc aac cac cat atc aaa gca tct gga tat tct gga aaa ata caa	2467
Tyr Ser Asn His His Ile Lys Ala Ser Gly Tyr Ser Gly Lys Ile Gln	
Tyr Ser Asn His His Ile Lys Ala Ser Gly Tyr Ser Gly Lys Ile Gln	
775 780 785	
acg gaa ggc aaa tgt tat agt acg aca tta ggg gcg gct ctc tct tgc	2515
Thr Glu Gly Lys Cys Tyr Ser Thr Thr Leu Gly Ala Ala Leu Ser Cys	
Thr Glu Gly Lys Cys Tyr Ser Thr Thr Leu Gly Ala Ala Leu Ser Cys	
790 795 800 805	
tct cta tct cta caa tgg cga tca cga cct ctc cac ttc act cct ttt	2563
Ser Leu Ser Leu Gln Trp Arg Ser Arg Pro Leu His Phe Thr Pro Phe	
Ser Leu Ser Leu Gln Trp Arg Ser Arg Pro Leu His Phe Thr Pro Phe	
810 815 820	
atc caa gca att gcc gtt cgt tct aat caa act gcg ttt caa gaa agt	2611
Ile Gln Ala Ile Ala Val Arg Ser Asn Gln Thr Ala Phe Gln Glu Ser	
Ile Gln Ala Ile Ala Val Arg Ser Asn Gln Thr Ala Phe Gln Glu Ser	
825 830 835	
gga gat aaa gct aga aaa ttt tct gtt cat aaa ccc tta tat aac ctg	2659
Gly Asp Lys Ala Arg Lys Phe Ser Val His Lys Pro Leu Tyr Asn Leu	
Gly Asp Lys Ala Arg Lys Phe Ser Val His Lys Pro Leu Tyr Asn Leu	
840 845 850	
aca gtc cct ctg gga att cag agc gct tgg gaa tcc aag ttc cgt ctt	2707
Thr Val Pro Leu Gly Ile Gln Ser Ala Trp Glu Ser Lys Phe Arg Leu	
Thr Val Pro Leu Gly Ile Gln Ser Ala Trp Glu Ser Lys Phe Arg Leu	
855 860 865	
cct acc tat tgg aac ata gag ctt gct tat cag cct gtc ctc tac caa	2755
Pro Thr Tyr Trp Asn Ile Glu Leu Ala Tyr Gln Pro Val Leu Tyr Gln	
Pro Thr Tyr Trp Asn Ile Glu Leu Ala Tyr Gln Pro Val Leu Tyr Gln	
870 875 880 885	
caa aat cct gag atc aac gtg agt cta gaa tct agt gga tcg tca tgg	2803
Gln Asn Pro Glu Ile Asn Val Ser Leu Glu Ser Ser Gly Ser Ser Trp	
Gln Asn Pro Glu Ile Asn Val Ser Leu Glu Ser Ser Gly Ser Ser Trp	
890 895 900	
ctc cta tca gga acc acc ctt gct cgc aat gcc att gct ttt aaa gga	2851
Leu Leu Ser Gly Thr Thr Leu Ala Arg Asn Ala Ile Ala Phe Lys Gly	
Leu Leu Ser Gly Thr Thr Leu Ala Arg Asn Ala Ile Ala Phe Lys Gly	
905 910 915	
aga aac caa att ttt atc ttc cct aaa ctt tcg gtg ttc tta gac tat	2899
Arg Asn Gln Ile Phe Ile Phe Pro Lys Leu Ser Val Phe Leu Asp Tyr	
Arg Asn Gln Ile Phe Ile Phe Pro Lys Leu Ser Val Phe Leu Asp Tyr	
920 925 930	
caa ggc tcg gta tcc tca tca acg acg aca cat tac ctt cac gca gga	2947
Gln Gly Ser Val Ser Ser Ser Thr Thr Thr His Tyr Leu His Ala Gly	
Gln Gly Ser Val Ser Ser Ser Thr Thr Thr His Tyr Leu His Ala Gly	
935 940 945	
acg acc ttt aag ttt taaaagcatg ttatatagac aatgcaacct gtaaagacca	3002
Thr Thr Phe Lys Phe	
Thr Thr Phe Lys Phe	
950	
aatagagagt agtgaacact ctctaccatc atgaatctta tgggagaagc taagggaat	3062
ccacagatac gtttccccc taaaaattaa gaaccgata catcctcact agagattcga	3122
aagaactact taaatcctaa gcattcga	3150

Fig. 19 (con't)

gtc	aca	gct	aaa	cat	att	aat	acg	gat	aat	ttc	tac	cct	gag	ggg	cta	1891
Val	Thr	Ala	Lys	His	Ile	Asn	Thr	Asp	Asn	Phe	Tyr	Pro	Glu	Gly	Leu	
Val	Thr	Ala	Lys	His	Ile	Asn	Thr	Asp	Asn	Phe	Tyr	Pro	Glu	Gly	Leu	
			585					590					595			
aat	aca	act	caa	cac	tac	ggc	tac	caa	ggc	gtt	tgg	tcc	cct	tac	tgg	1939
Asn	Thr	Thr	Gln	His	Tyr	Gly	Tyr	Gln	Gly	Val	Trp	Ser	Pro	Tyr	Trp	
Asn	Thr	Thr	Gln	His	Tyr	Gly	Tyr	Gln	Gly	Val	Trp	Ser	Pro	Tyr	Trp	
		600					605					610				
atc	gaa	aca	atc	aca	act	tct	gat	acc	tct	tct	gaa	gat	act	gtg	aat	1987
Ile	Glu	Thr	Ile	Thr	Thr	Ser	Asp	Thr	Ser	Ser	Glu	Asp	Thr	Val	Asn	
Ile	Glu	Thr	Ile	Thr	Thr	Ser	Asp	Thr	Ser	Ser	Glu	Asp	Thr	Val	Asn	
		615				620					625					
act	tta	cat	cgc	cag	ctt	tat	ggg	gat	tgg	aca	cct	aca	gga	tat	aag	2035
Thr	Leu	His	Arg	Gln	Leu	Tyr	Gly	Asp	Trp	Thr	Pro	Thr	Gly	Tyr	Lys	
Thr	Leu	His	Arg	Gln	Leu	Tyr	Gly	Asp	Trp	Thr	Pro	Thr	Gly	Tyr	Lys	
		630			635					640					645	
gta	aac	cca	gaa	aac	aaa	gga	gac	att	gcc	cta	tct	gcc	ttc	tgg	caa	2083
Val	Asn	Pro	Glu	Asn	Lys	Gly	Asp	Ile	Ala	Leu	Ser	Ala	Phe	Trp	Gln	
Val	Asn	Pro	Glu	Asn	Lys	Gly	Asp	Ile	Ala	Leu	Ser	Ala	Phe	Trp	Gln	
				650					655					660		
tct	ttc	cat	aac	tta	ttt	gcg	aca	cta	cgt	tat	caa	aca	cag	caa	ggc	2131
Ser	Phe	His	Asn	Leu	Phe	Ala	Thr	Leu	Arg	Tyr	Gln	Thr	Gln	Gln	Gly	
Ser	Phe	His	Asn	Leu	Phe	Ala	Thr	Leu	Arg	Tyr	Gln	Thr	Gln	Gln	Gly	
			665					670					675			
caa	ata	gca	cct	aca	gct	tct	gga	gaa	gct	act	cga	ctc	ttc	gtg	cat	2179
Gln	Ile	Ala	Pro	Thr	Ala	Ser	Gly	Glu	Ala	Thr	Arg	Leu	Phe	Val	His	
Gln	Ile	Ala	Pro	Thr	Ala	Ser	Gly	Glu	Ala	Thr	Arg	Leu	Phe	Val	His	
		680					685					690				
caa	aat	agc	aac	aat	gat	gcg	aaa	gga	ttc	cat	atg	gaa	gct	acg	ggg	2227
Gln	Asn	Ser	Asn	Asn	Asp	Ala	Lys	Gly	Phe	His	Met	Glu	Ala	Thr	Gly	
Gln	Asn	Ser	Asn	Asn	Asp	Ala	Lys	Gly	Phe	His	Met	Glu	Ala	Thr	Gly	
		695				700					705					
tat	tct	ttg	gga	aca	acc	tca	aac	act	gct	tct	aat	cat	agc	ttt	ggg	2275
Tyr	Ser	Leu	Gly	Thr	Thr	Ser	Asn	Thr	Ala	Ser	Asn	His	Ser	Phe	Gly	
Tyr	Ser	Leu	Gly	Thr	Thr	Ser	Asn	Thr	Ala	Ser	Asn	His	Ser	Phe	Gly	
		710			715				720						725	
gta	aac	ttc	tcc	caa	ctt	ttc	agt	aat	ctc	tac	gag	agc	cac	tcc	gac	2323
Val	Asn	Phe	Ser	Gln	Leu	Phe	Ser	Asn	Leu	Tyr	Glu	Ser	His	Ser	Asp	
Val	Asn	Phe	Ser	Gln	Leu	Phe	Ser	Asn	Leu	Tyr	Glu	Ser	His	Ser	Asp	
				730					735					740		
aat	tcc	gtg	gct	tcg	cat	acg	aca	act	gta	gcg	ctc	cag	atc	aat	aat	2371
Asn	Ser	Val	Ala	Ser	His	Thr	Thr	Thr	Val	Ala	Leu	Gln	Ile	Asn	Asn	
Asn	Ser	Val	Ala	Ser	His	Thr	Thr	Thr	Val	Ala	Leu	Gln	Ile	Asn	Asn	
			745					750					755			
cct	tgg	ctg	caa	gag	aga	ttc	tct	aca	tct	gca	tct	cta	gcc	tac	agc	2419
Pro	Trp	Leu	Gln	Glu	Arg	Phe	Ser	Thr	Ser	Ala	Ser	Leu	Ala	Tyr	Ser	
Pro	Trp	Leu	Gln	Glu	Arg	Phe	Ser	Thr	Ser	Ala	Ser	Leu	Ala	Tyr	Ser	
		760				765						770				

Figure 20 (RY-44)

## Restriction enzyme analysis of CPN100622

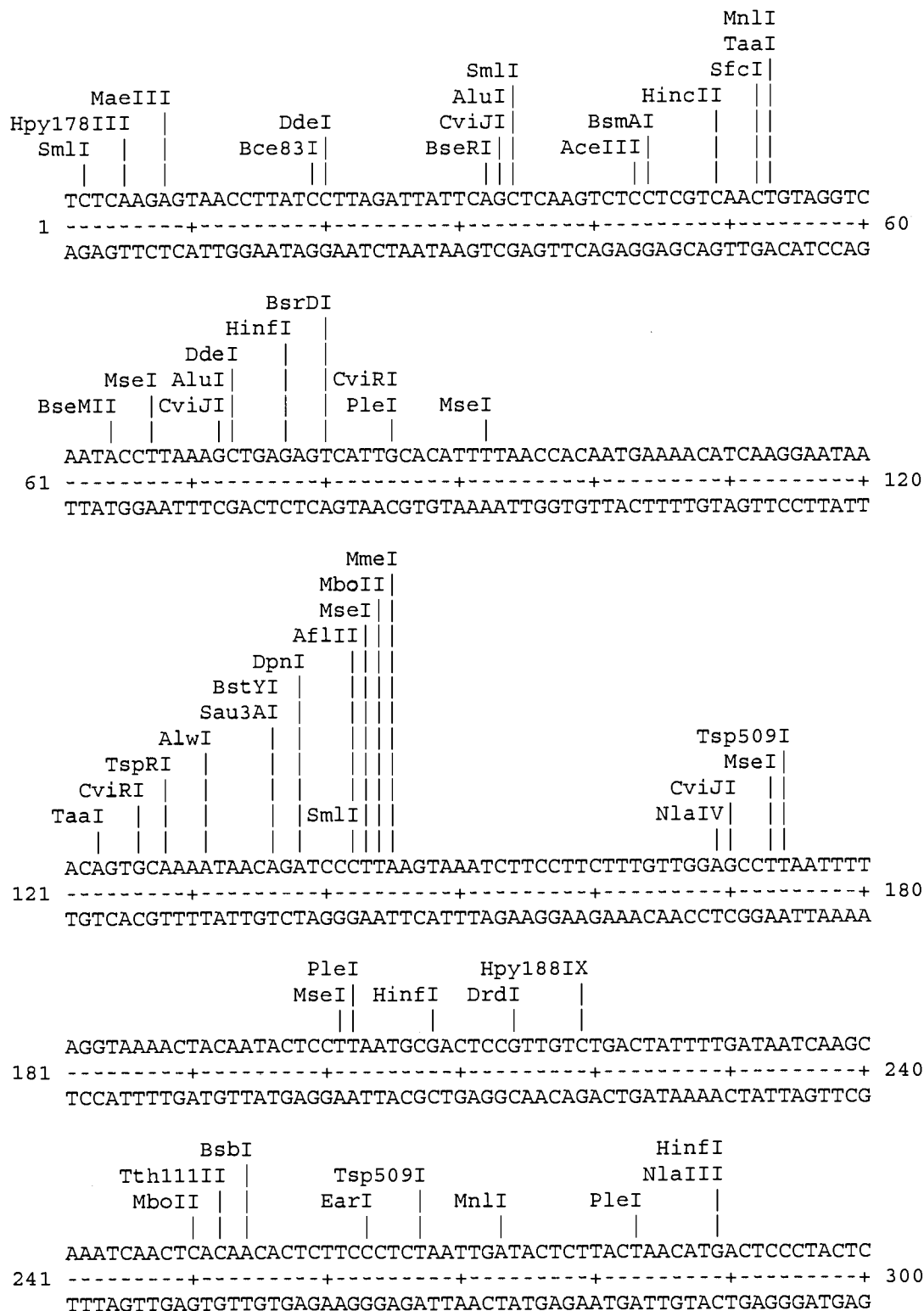
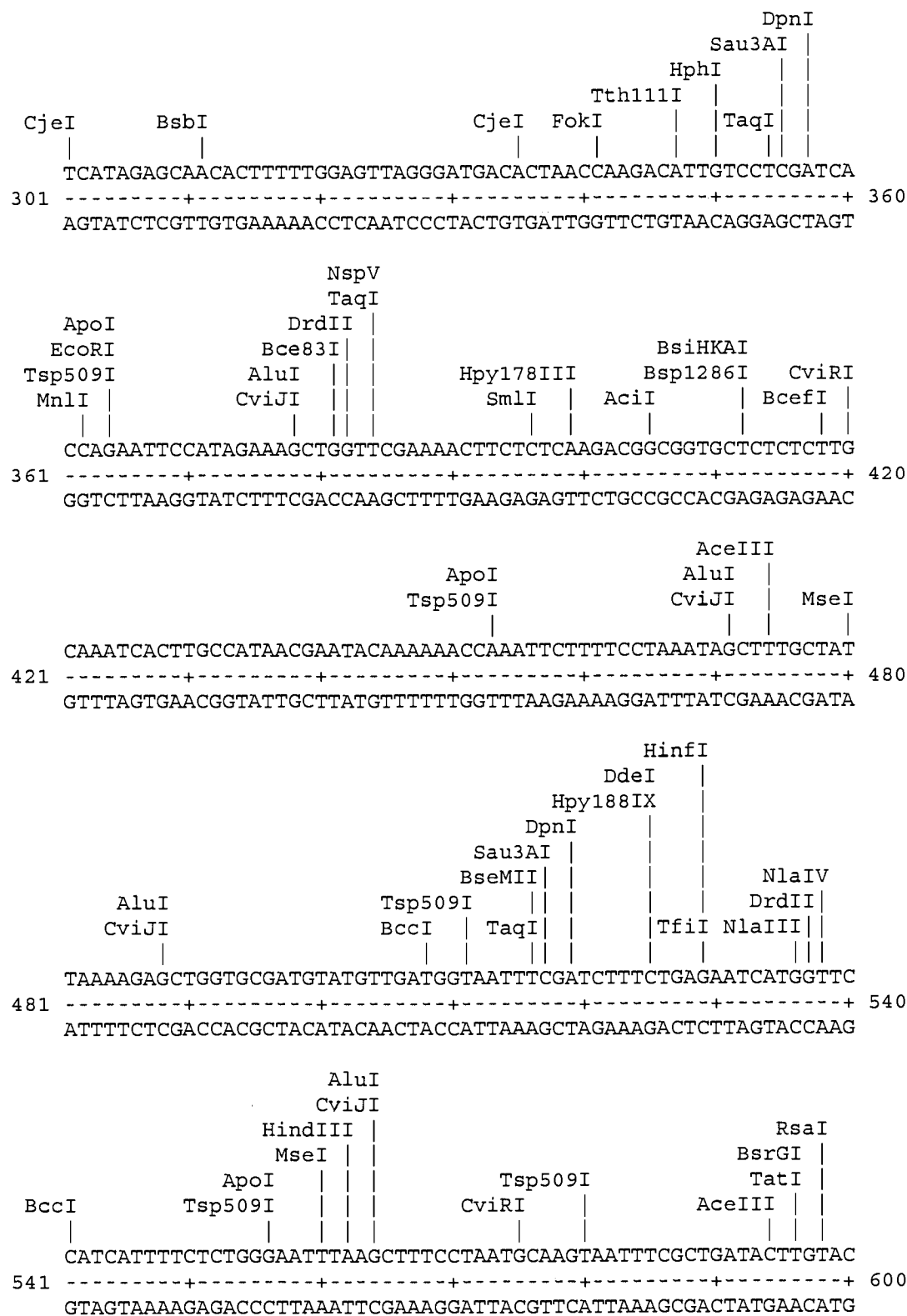


Fig. 20 (con't)



```

      AluI      NspV      MaeIII      AciI
      CviJI      TaqI      NlaIV
      |          |          |          |
AGGGGGAGCTGTTTTATGTTTCGAAAAATGTTACAATCTCAAAAAATCAAGGAACCGCATA
601 -----+-----+-----+-----+-----+-----+-----+ 660
TCCCCCTCGACAAAATACAAGCTTTTTACAATGTTAGAGTTTTTTAGTTCCTTGGCGTAT

      CviRI
      BseRI
      Fnu4HI
      AluI
      CviJI
      TseI
      Eco57I      Hpy178III      BbvI      Hin4I
      MseI      MboII      MnlI      MwoI      MseI
      |          |          |          |          |          |
CTTCATTAACAACAAGGCAAAATCTTCAGGAGGAGCAATCCAAGCTGCAATCATAAACAT
661 -----+-----+-----+-----+-----+-----+-----+ 720
GAAGTAATTGTTGTTCCGTTTTAGAAGTCCTCCTCGTTAGGTTTCGACGTTAGTATTTGTA

      EciI
      FauI
      AciI
      Sth132I
      Cac8I
      PstI
      BsrI
      TspRI
      CviJI
      HaeIII
      Sau96I
      BsbI
      BbvI
      MseI
      TseI
      Fnu4HI
      SfcI
      MspA1I
      MwoI
      AciI
      |          |          |          |          |          |
TAAGGACAACACTGGCCCTTGCCTGTTTTTTAATAATGCTGCAGGCGGAACAGCGGGGGG
721 -----+-----+-----+-----+-----+-----+-----+ 780
ATTCTGTTGTGACCGGGAACGGACAAAAAATTATTACGACGTCCGCCTTGTCGCCCCC

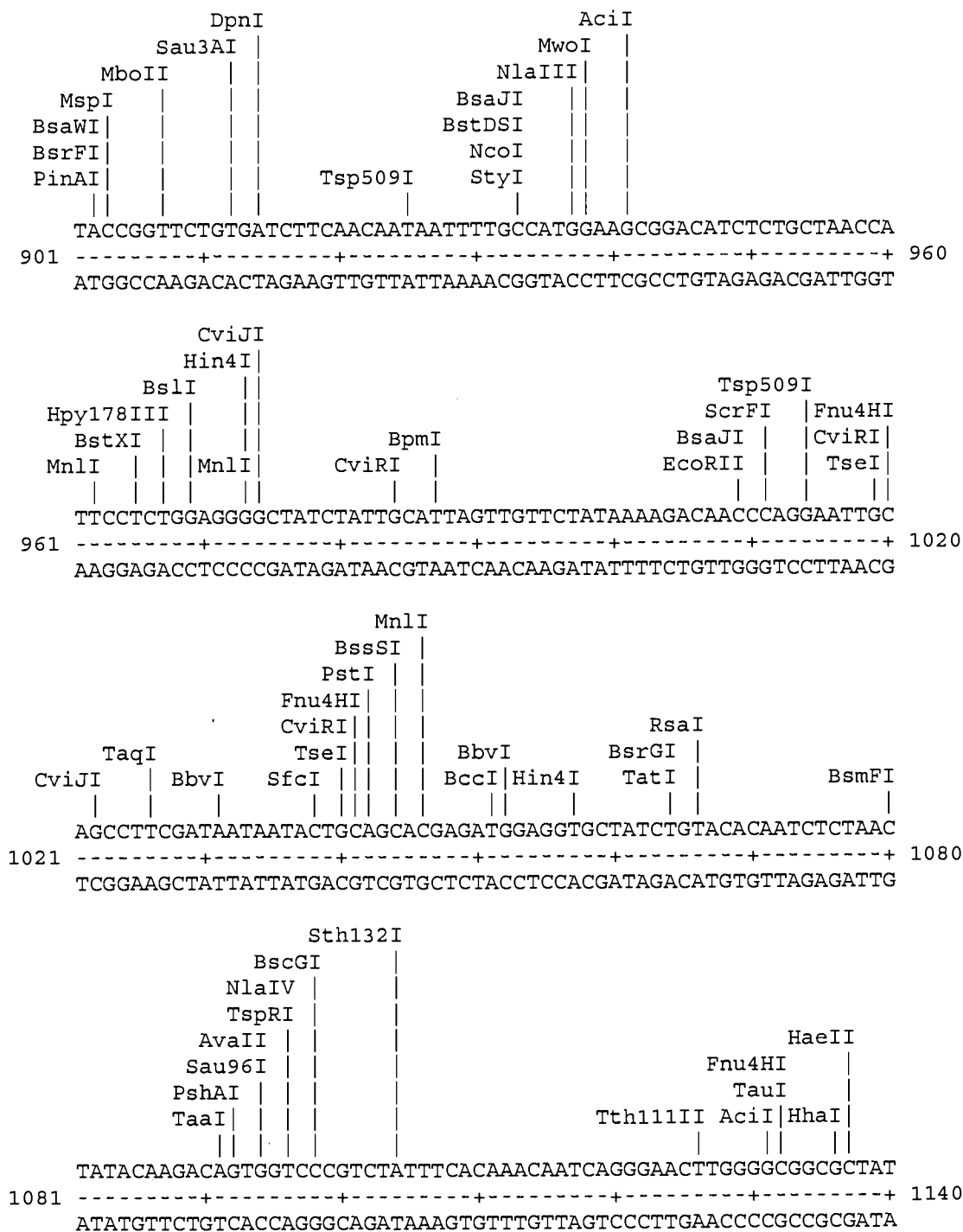
      MwoI
      Tth111III
      HhaI
      ThaI
      Tsp509I
      Tsp509I
      DdeI
      CviJI
      BseMII
      |          |          |          |          |          |
CGCGTTGTTTCGCTAATGCTTGTAGAATTGAGAATAATTCTCAGCCTATCTATTTTTTGAA
781 -----+-----+-----+-----+-----+-----+-----+ 840
GCGCAACAAGCGATTACGAACATCTTAACTCTTATTAAGAGTCGGATAGATAAAAAACTT

      Hpy178III
      RsaI
      BsmI
      TaqII
      CviRI
      TatI
      CviRI
      |          |          |          |
TAACCAATCAGGTCTGGGTGGTGCAATAAGAGTACATCAAGAGTGCATTCTTACAAAGAA
841 -----+-----+-----+-----+-----+-----+-----+ 900
ATTGGTTAGTCCAGACCCACCACGTTATTCTCATGTAGTTCTCACGTAAGAATGTTTCTT

```



Fig. 20 (con't)



CviRI  
 NlaIII  
 NspI  
 SphI  
 Cac8I  
 DpnI  
 Hpy178III  
 CviRI  
 BclI  
 NlaIII  
 BccI  
 Sau3AI  
 CATGCTCCGTCAAGATGGTGCATGCACTTTATTTTGCTGATCAGGGAGATATTATTTTTTA  
 1141 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 1200  
 GTACGAGGCAGTTCTACCACGTACGTGAAATAAACGACTAGTCCCTCTATAATAAAAAAT

MmeI  
 ThaI  
 AflIII  
 Cac8I  
 MluI  
 CviRI  
 NlaIII  
 BsgI  
 TAATAATAGACACTTCAAAGATACTTTTCAGCAATCATGTTTCTGTAAACTGCACGCGTAA  
 1201 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 1260  
 ATTATTATCTGTGAAGTTTCTATGAAAGTCGTTAGTACAAAGACATTTGACGTGCGCATT

DpnI  
 Sau3AI  
 AlwI  
 MseI  
 BsmAI  
 TaaI  
 TGTCTCATTAACAGTTGGAGCAAGTCAAGGTCATTCTGCTACCTTCTATGATCCCATACT  
 1261 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 1320  
 ACAGAGTAATTGTCAACCTCGTTCAGTTCAGTAAGACGATGGAAGATACTAGGGTATGA

CjePI  
 MseI  
 ApoI  
 Tsp509I  
 Hpy178III  
 BsaJI  
 Hpy188IX  
 ACAAAGATATACTATACAAAACCTCTATCCAAAATTTAATCCTAATCCAGAACACCTCGG  
 1321 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 1380  
 TGTTTCTATATGATATGTTTTGAGATAGGTTTTTAAATTAGGATTAGGTCTTGTGGAGCC

Hpy178III  
 MspI  
 BsaWI  
 BspEI  
 BciVI  
 MnII  
 BseRI  
 CjePI  
 TaqI  
 BssSI  
 Hpy178III  
 AACTATCTTGTTCCTCAACATATATCCGGATACATCGACTTCTCGTGATGACTTCAT  
 1381 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 1440  
 TTGATAGAACAAGAGGAGTTGTATATAAGGCCATGTAGCTGAAGAGCACTACTGAAGTA

Fig. 20 (con't)

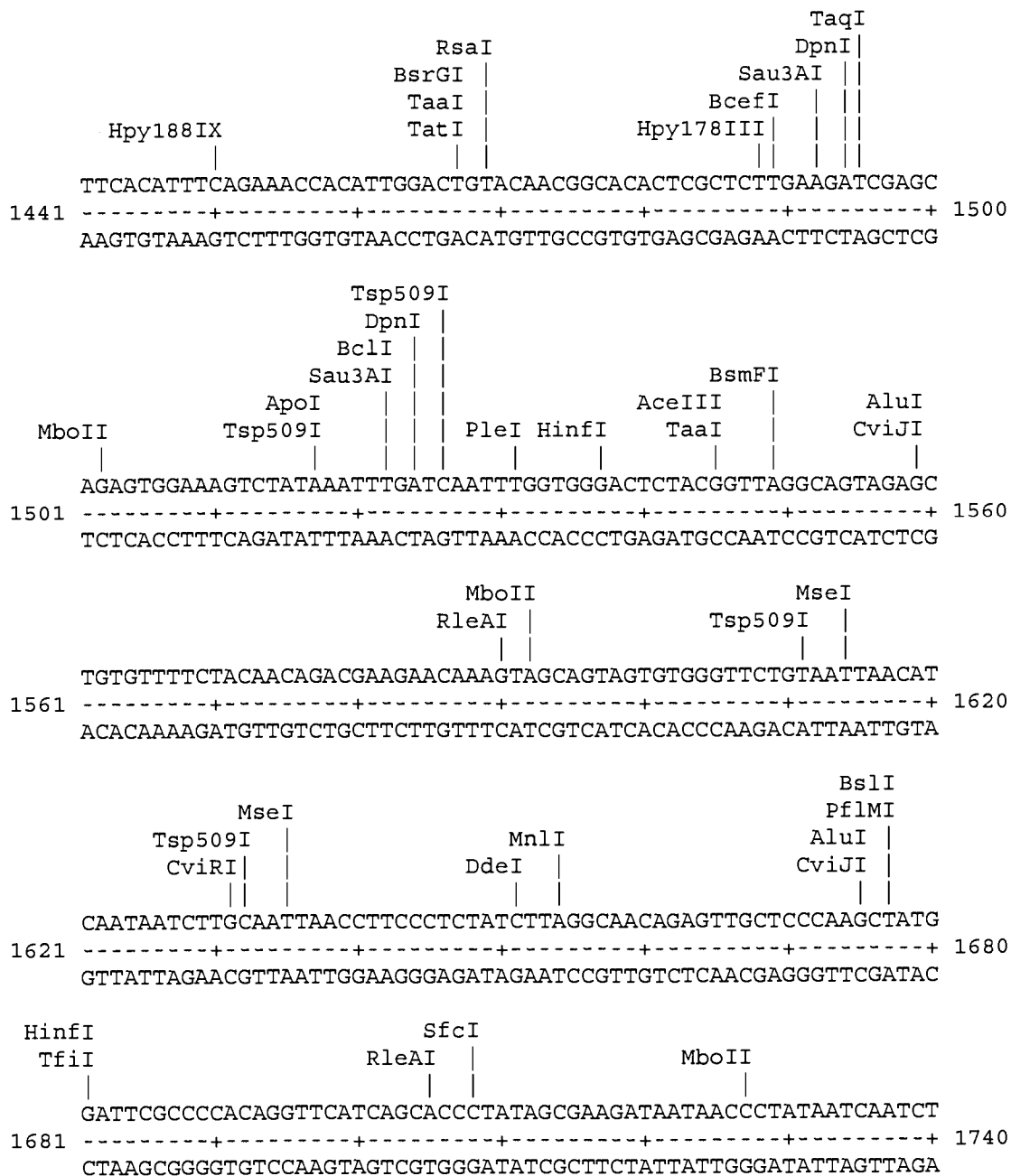


Fig. 20 (cont)

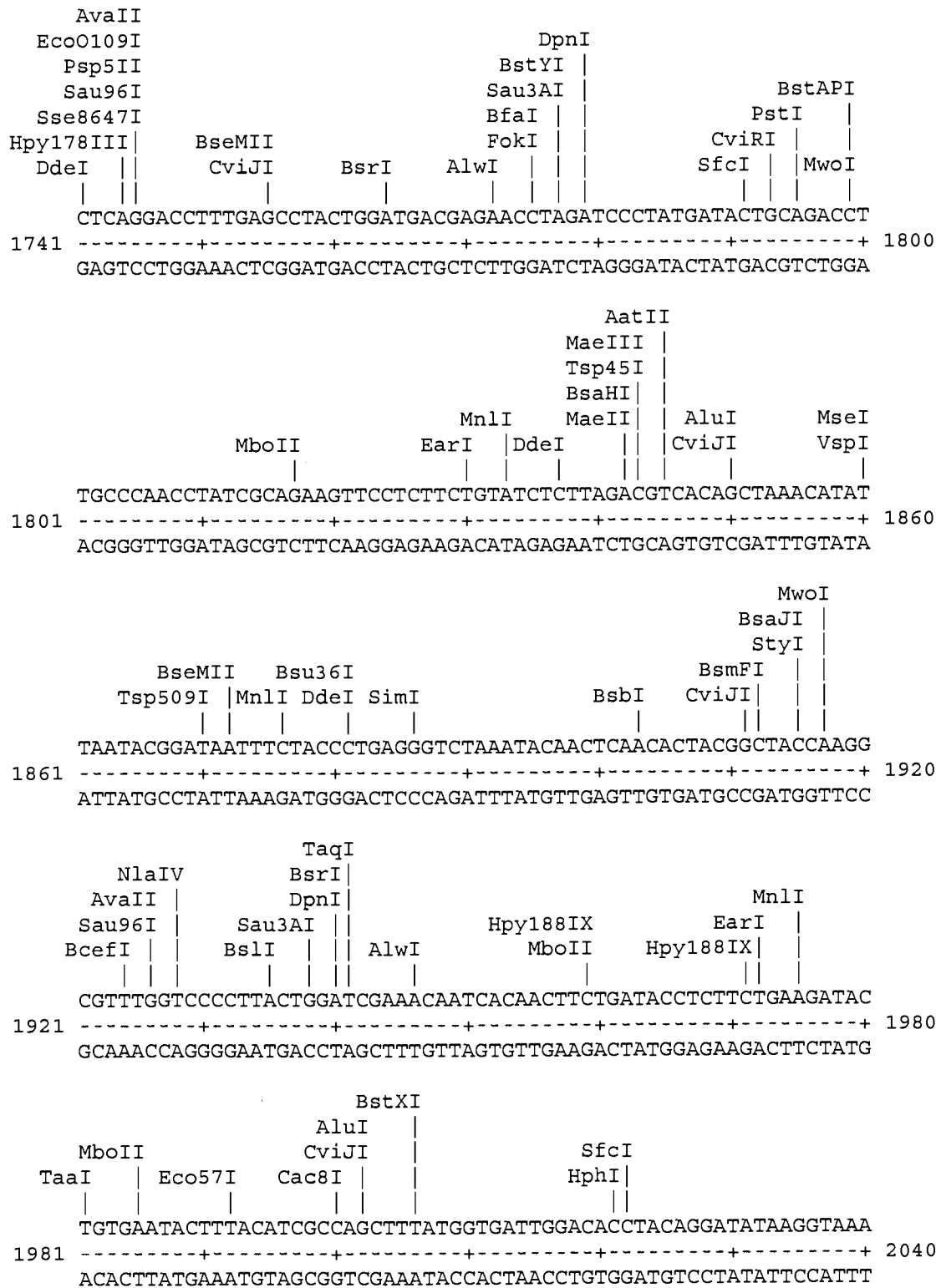


Fig. 20 (con't)

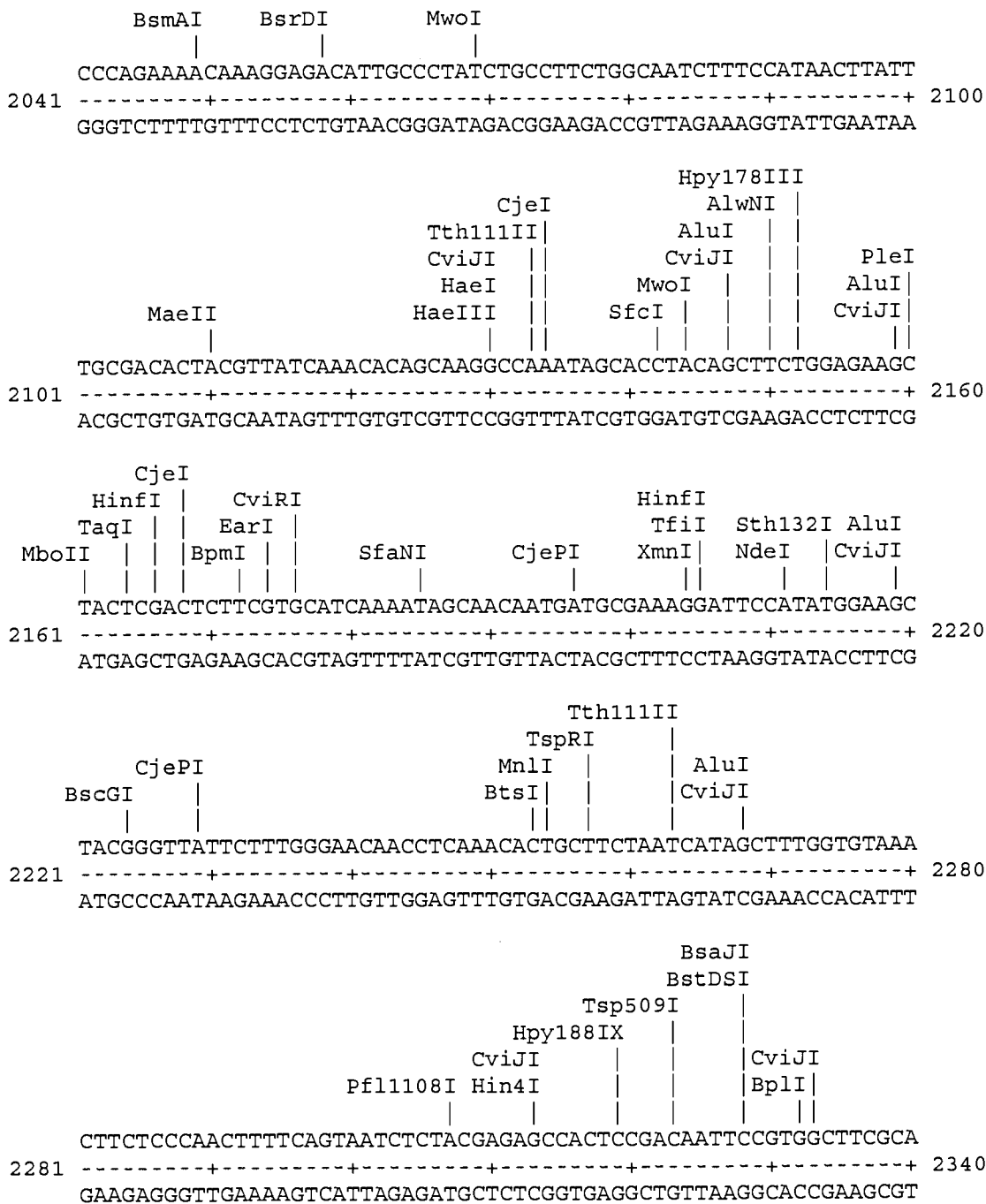


Fig. 20 (cont)

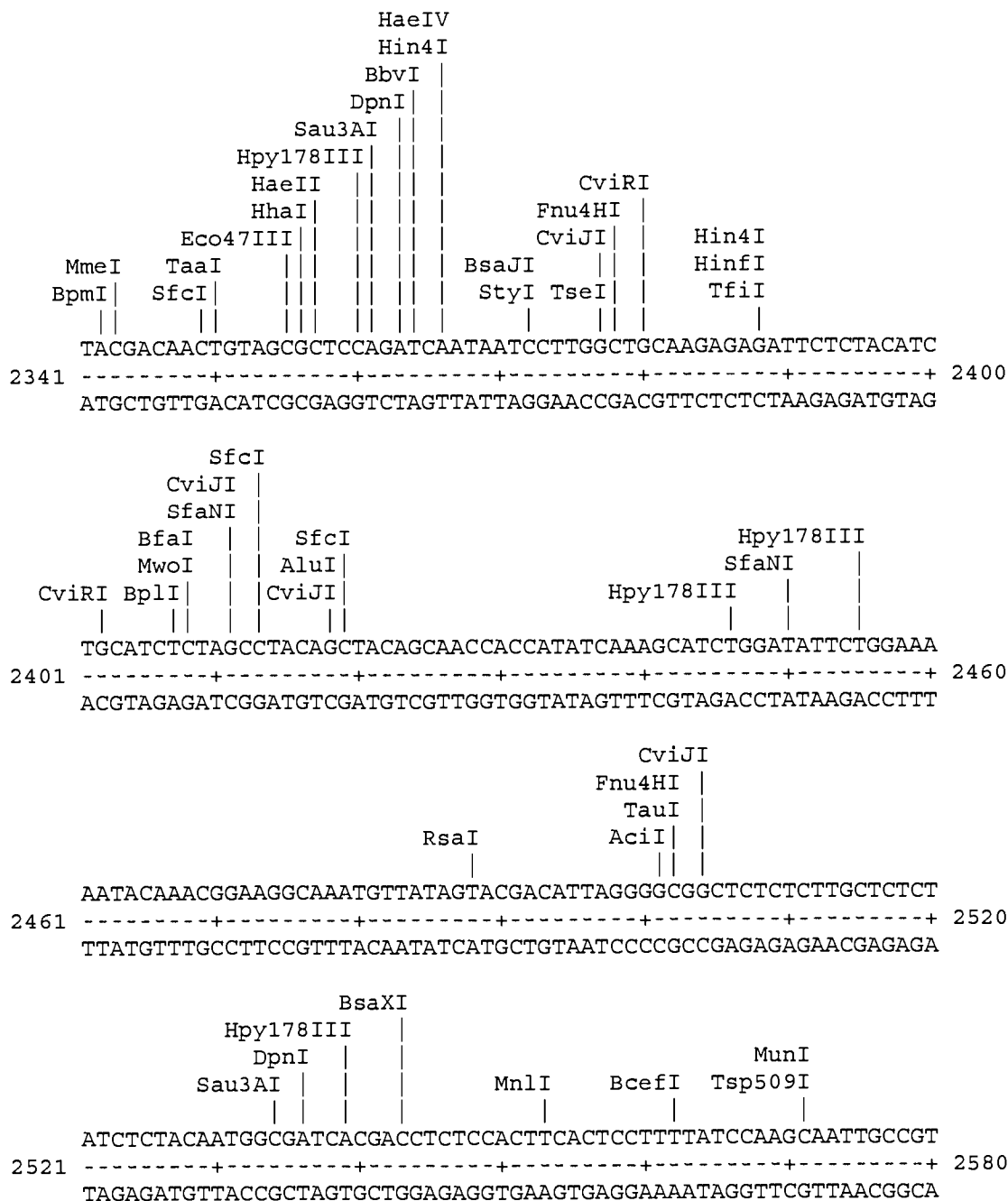


Fig. 20 (con't)

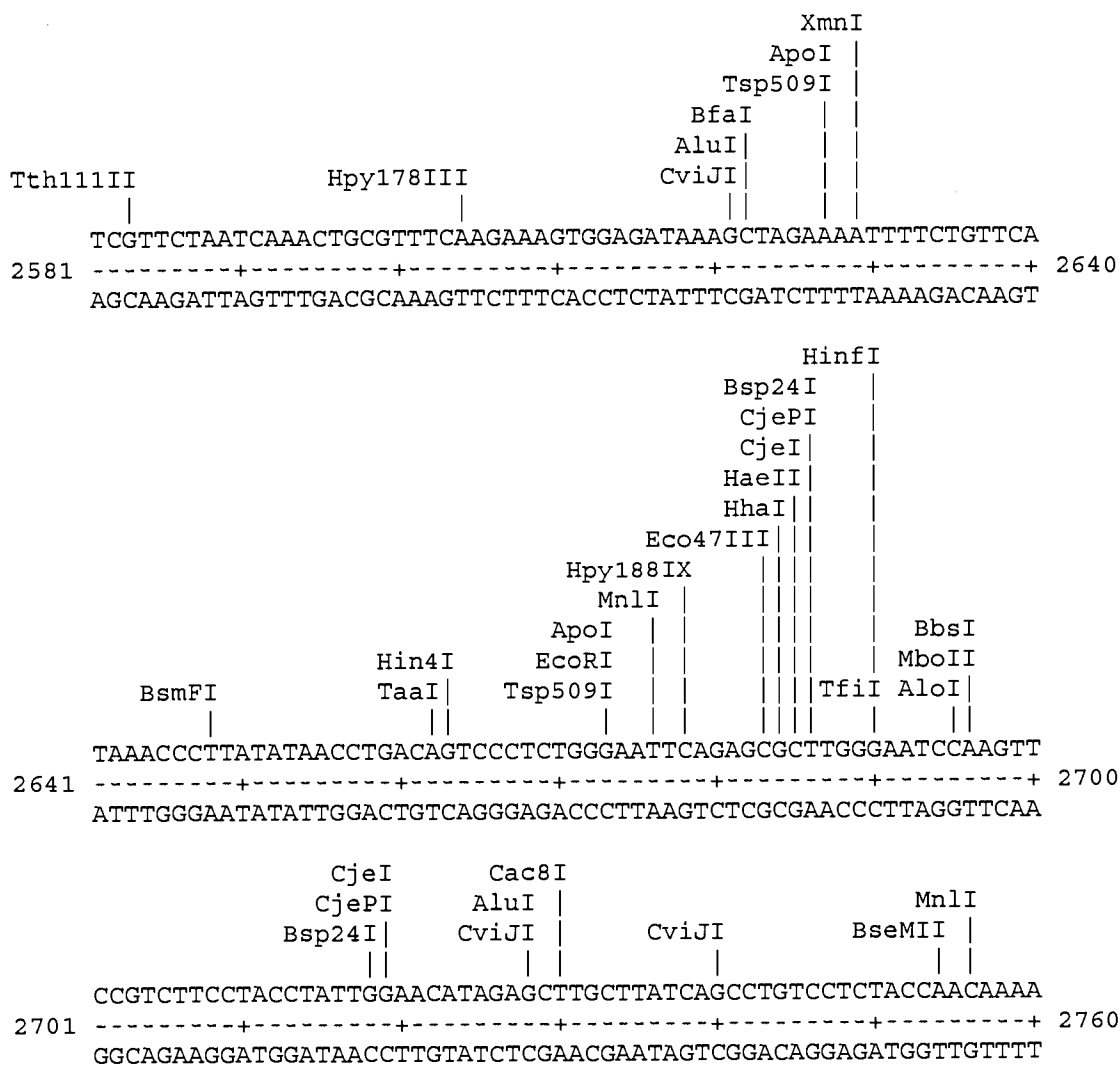


Fig. 20 (con't)

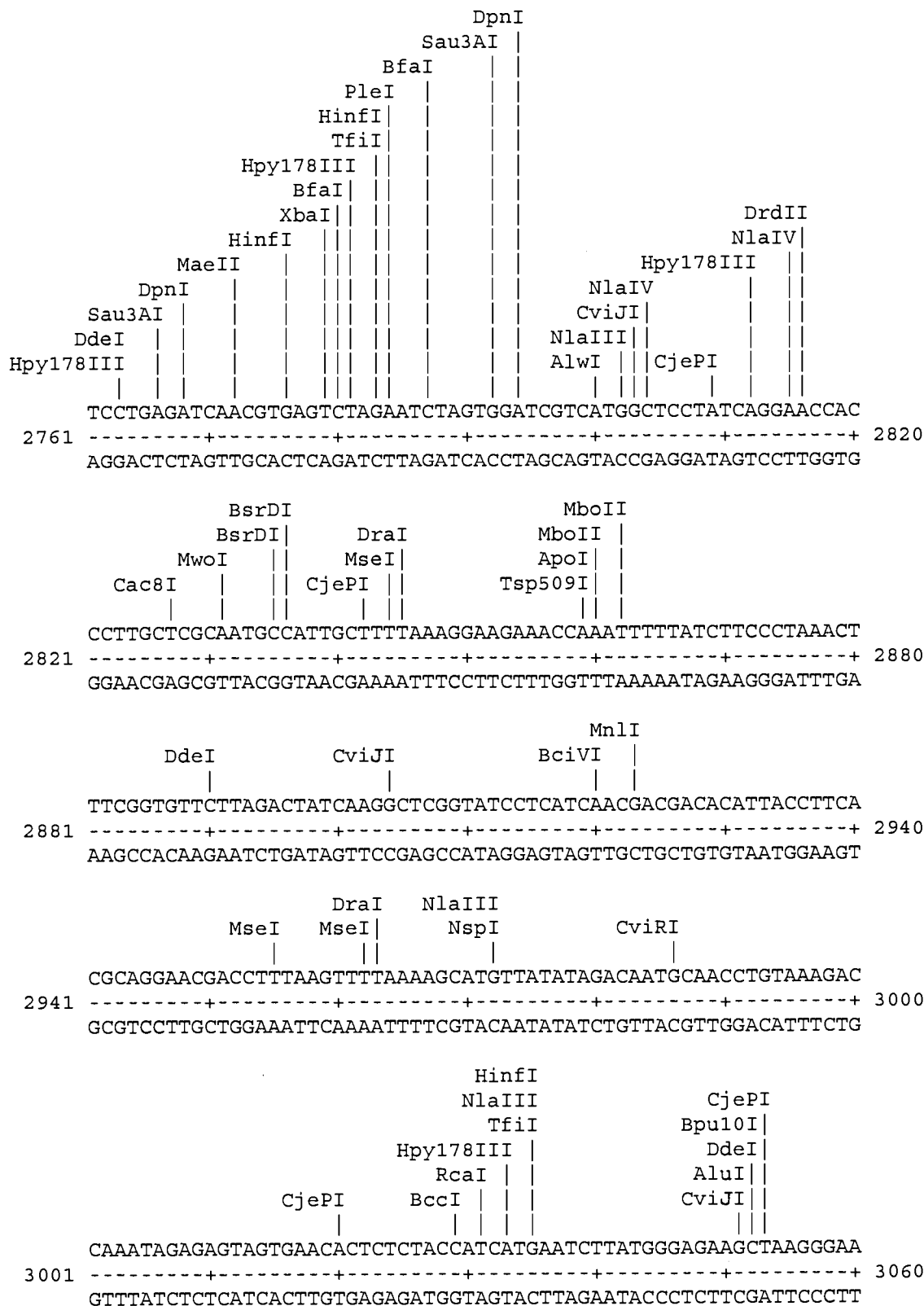




Fig. 20 (con't)

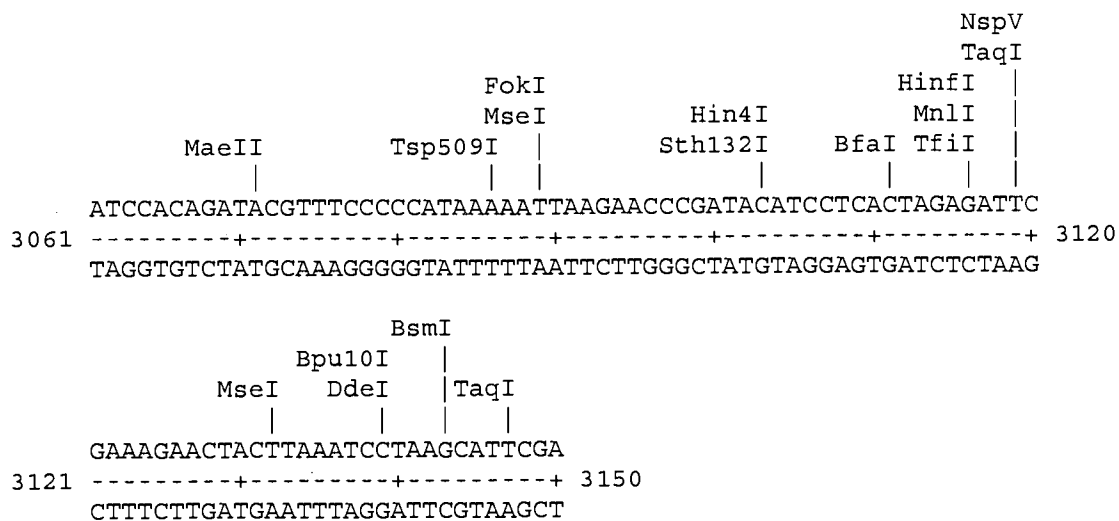


Figure 21A: CPN100626 Coding Sequence

tccctgaactc	cactcgaaat	tactgattag	ccaaggtacg	tggacgacgc	aggccactcc	60
tgtgacctac	aatgcttttag	ggatcaaagt	gaaaaatacc	atgcaggtgt	ttcctaaagt	120
cactctctcc	ttagattact	ctgcggatat	ttcttcctcc	acgctgagtc	actacttaaa	180
cgtggcgagt	agaatgagat	ttttaacaat	aagtgaccaa	aacagaaaga	ttaaggaacc	240
tctagtgtca	aagactcctc	ctaagttttt	attctatctc	gggaatttca	cagcctgcat	300
gttcgggatg	actcctgcag	tgtatagttt	acaaacggac	tcccttgaaa	agtttgcttt	360
agagagggat	gaagagtttc	gtacgagctt	tcctctctta	gactctctct	ccactcttac	420
aggattttct	ccaataacta	cgtttgttgg	aaatagacat	aattcctctc	aagacattgt	480
acttttctaac	tacaagtcta	ttgataacat	ccttcttctt	tggacatcgg	ctggggggagc	540
tgtgtcctgt	aataattttct	tattatcaaa	tgttgaagac	catgccttct	tcagtaaaaa	600
tctcgcgatt	gggactggag	gcgcgattgc	ttgccaggga	gcctgcacaa	tcacgaagaa	660
tagaggacct	cttatttttt	tcagcaatcg	aggtcttaac	aatgcgagta	caggaggaga	720
aactcgtggg	ggtgcgattg	cctgtaatgg	agacttcacg	atttctcaaa	atcaagggac	780
tttctacttt	gtcaacaatt	ccgtcaacaa	ctggggagga	gccctctcca	ccaatggaca	840
ctgccgcctc	caaagcaaca	gggcacctct	actctttttt	aacaatacag	cccctagtgg	900
agggggtgcg	cttcgtagtg	aaaatacaac	gatctctgat	aacacgcgtc	ctattttattt	960
taagaacaac	tgtgggaaca	atggcggggc	cattcaaaaca	agcgttactg	ttgcgataaa	1020
aaataactcc	gggtcggtga	ttttcaataa	caacacagcg	ttatctgggt	cgataaaattc	1080
aggaaatggt	tcaggagggg	cgattttatac	aacaaacctt	tccatagacg	ataaccctgg	1140
aactattctt	ttcaataata	actactgcat	tcgcgatggc	ggagctatct	gtacacaatt	1200
tttgacaatc	aaaaatagtg	gccacgtata	tttcaccaac	aatcaaggaa	actggggagg	1260
tgctcttatg	ctctacaggg	acagcacctg	cctactcttc	gcggaacaag	gaaatatcgc	1320
atttcaaaat	aatgaggttt	tcctcaccac	atttggtaga	tacaacgcca	tacattgtac	1380
accaaatagc	aacttacaac	ttggagctaa	taaggggtat	acgactgctt	tttttgatcc	1440
tatagaacac	caacatccaa	ctacaaatcc	tctaattctt	aatcccaatg	cgaaccatca	1500
gggaacgata	ttattttctt	cagcctatat	cccagaagct	tctgactacg	aaaataattt	1560
cattagcagc	tcgaaaaata	cctctgaact	tcgcaatggg	gtcctctcta	tcgaggatcg	1620
tgcgggatgg	caattctata	agttcactca	aaaaggaggt	atccttaa	tagggcatgc	1680
ggcgagtatt	gcaacaactg	ccaactctga	gactccatca	actagtgtag	gctcccagg	1740
catcattaat	aaccttgcca	ttaacctccc	ctcgatctta	gcaaaaagg	aaagctcctac	1800
cttgtggatc	cgctctctac	aatctagtgc	tcctttcaca	gaggacaata	accctacaat	1860
tacttttatca	ggtcctctga	cactcttaaa	tgaggaaaac	cgcgatccct	acgacagtat	1920
agatctctct	gagcctttac	aaaacattga	tcttctttct	ttatcggatg	taacagcacg	1980
tcatatcaat	accgataact	ttcatcctga	aagcttaaat	gcgactgagc	attacggtta	2040
tcaaggcatc	tggctctcct	attgggtaga	gacgataaca	acaacaaata	acgcttctat	2100
agagacggca	aacaccctct	acagagctct	gtatgccaat	tggactccct	taggatataa	2160
ggtcaatcct	gaataccaag	gagatcttgc	tacgactccc	ctatggcaat	cctttcatac	2220
tatgttctct	ctattaagaa	gttataatcg	aactgggtgat	tctgatatcg	agaggccttt	2280
cttagaaaatt	caagggattg	ccgacggcct	ctttgttcat	caaaatagca	tccccggggc	2340
tccaggattc	cgtatccaat	ctacagggtta	ttccttacaa	gcactcctcg	aaacttcttt	2400
acatcagaaa	atctccttag	gttttgcaca	gttcttcacc	cgactaaag	aaatcggatc	2460
aagcaacaac	gtctcggctc	acaatacagt	ctcttcactt	tatgttgagc	ttccgtgggt	2520
ccaagaggcc	tttgcaacat	cccacagttt	agcgtatggc	tatggggacc	atcacctcca	2580
cgctacatc	cgtcacatca	agaacagggc	agaagggacg	tgttatagcc	atacattagc	2640
agcagctatc	ggctgttctt	tcccttggca	acagaaatcc	tatcttcacc	tcagcccgtt	2700
cgttcaggca	attgcaatac	gttctcacca	aacagcggtc	gaagagattg	gtgacaatcc	2760
ccgaaagttt	gtctctcaaa	agcctttcta	taatctgacc	ttaccttag	gaatccaagg	2820
aaaatggcag	tcaaaattcc	acgtacctac	agaatggact	ctagaacttt	cttaccaacc	2880
ggtactctat	caacaaaatc	cccaaactcg	tgtcacgcta	cttgcgagcg	gaggttcctg	2940
ggatatccta	ggccataact	atgttcgcaa	tgctttaggg	tacaaaagtc	acaatcaaac	3000
tgcgctcttc	cgttctctcg	atctattctt	ggattaccaa	ggatcggtct	cctcctcgac	3060
atctacgcac	catctccaag	caggaagtac	cttaaaattc	taaaataaaa	gaacgataaa	3120
attgaaatct	ttagaattaa	caactatccg	atgagctacg	ttagcccaat	cggtagagga	3180
ctccctcaaa	atttaataaa					3200

Figure 21B: CPN100626 Deduced Amino Acid Sequence

Met	Gln	Val	Phe	Pro	Lys	Val	Thr	Leu	Ser	Leu	Asp	Tyr	Ser	Ala	Asp	1	5	10	15
Ile	Ser	Ser	Ser	Thr	Leu	Ser	His	Tyr	Leu	Asn	Val	Ala	Ser	Arg	Met	20	25	30	
Arg	Phe	Leu	Thr	Ile	Ser	Asp	Gln	Asn	Arg	Lys	Ile	Lys	Glu	Pro	Leu	35	40	45	
Val	Ser	Lys	Thr	Pro	Pro	Lys	Phe	Leu	Phe	Tyr	Leu	Gly	Asn	Phe	Thr	50	55	60	
Ala	Cys	Met	Phe	Gly	Met	Thr	Pro	Ala	Val	Tyr	Ser	Leu	Gln	Thr	Asp	65	70	75	80
Ser	Leu	Glu	Lys	Phe	Ala	Leu	Glu	Arg	Asp	Glu	Glu	Phe	Arg	Thr	Ser	85	90	95	
Phe	Pro	Leu	Leu	Asp	Ser	Leu	Ser	Thr	Leu	Thr	Gly	Phe	Ser	Pro	Ile	100	105	110	
Thr	Thr	Phe	Val	Gly	Asn	Arg	His	Asn	Ser	Ser	Gln	Asp	Ile	Val	Leu	115	120	125	
Ser	Asn	Tyr	Lys	Ser	Ile	Asp	Asn	Ile	Leu	Leu	Leu	Trp	Thr	Ser	Ala	130	135	140	
Gly	Gly	Ala	Val	Ser	Cys	Asn	Asn	Phe	Leu	Leu	Ser	Asn	Val	Glu	Asp	145	150	155	160
His	Ala	Phe	Phe	Ser	Lys	Asn	Leu	Ala	Ile	Gly	Thr	Gly	Gly	Ala	Ile	165	170	175	
Ala	Cys	Gln	Gly	Ala	Cys	Thr	Ile	Thr	Lys	Asn	Arg	Gly	Pro	Leu	Ile	180	185	190	
Phe	Phe	Ser	Asn	Arg	Gly	Leu	Asn	Asn	Ala	Ser	Thr	Gly	Gly	Glu	Thr	195	200	205	
Arg	Gly	Gly	Ala	Ile	Ala	Cys	Asn	Gly	Asp	Phe	Thr	Ile	Ser	Gln	Asn	210	215	220	
Gln	Gly	Thr	Phe	Tyr	Phe	Val	Asn	Asn	Ser	Val	Asn	Asn	Trp	Gly	Gly	225	230	235	240
Ala	Leu	Ser	Thr	Asn	Gly	His	Cys	Arg	Ile	Gln	Ser	Asn	Arg	Ala	Pro	245	250	255	
Leu	Leu	Phe	Phe	Asn	Asn	Thr	Ala	Pro	Ser	Gly	Gly	Gly	Ala	Leu	Arg	260	265	270	
Ser	Glu	Asn	Thr	Thr	Ile	Ser	Asp	Asn	Thr	Arg	Pro	Ile	Tyr	Phe	Lys	275	280	285	

Fig. 21B (con't)

Asn	Asn	Cys	Gly	Asn	Asn	Gly	Gly	Ala	Ile	Gln	Thr	Ser	Val	Thr	Val	290	295	300	
Ala	Ile	Lys	Asn	Asn	Ser	Gly	Ser	Val	Ile	Phe	Asn	Asn	Asn	Thr	Ala	305	310	315	320
Leu	Ser	Gly	Ser	Ile	Asn	Ser	Gly	Asn	Gly	Ser	Gly	Gly	Ala	Ile	Tyr	325	330	335	
Thr	Thr	Asn	Leu	Ser	Ile	Asp	Asp	Asn	Pro	Gly	Thr	Ile	Leu	Phe	Asn	340	345	350	
Asn	Asn	Tyr	Cys	Ile	Arg	Asp	Gly	Gly	Ala	Ile	Cys	Thr	Gln	Phe	Leu	355	360	365	
Thr	Ile	Lys	Asn	Ser	Gly	His	Val	Tyr	Phe	Thr	Asn	Asn	Gln	Gly	Asn	370	375	380	
Trp	Gly	Gly	Ala	Leu	Met	Leu	Leu	Gln	Asp	Ser	Thr	Cys	Leu	Leu	Phe	385	390	395	400
Ala	Glu	Gln	Gly	Asn	Ile	Ala	Phe	Gln	Asn	Asn	Glu	Val	Phe	Leu	Thr	405	410	415	
Thr	Phe	Gly	Arg	Tyr	Asn	Ala	Ile	His	Cys	Thr	Pro	Asn	Ser	Asn	Leu	420	425	430	
Gln	Leu	Gly	Ala	Asn	Lys	Gly	Tyr	Thr	Thr	Ala	Phe	Phe	Asp	Pro	Ile	435	440	445	
Glu	His	Gln	His	Pro	Thr	Thr	Asn	Pro	Leu	Ile	Phe	Asn	Pro	Asn	Ala	450	455	460	
Asn	His	Gln	Gly	Thr	Ile	Leu	Phe	Ser	Ser	Ala	Tyr	Ile	Pro	Glu	Ala	465	470	475	480
Ser	Asp	Tyr	Glu	Asn	Asn	Phe	Ile	Ser	Ser	Ser	Lys	Asn	Thr	Ser	Glu	485	490	495	
Leu	Arg	Asn	Gly	Val	Leu	Ser	Ile	Glu	Asp	Arg	Ala	Gly	Trp	Gln	Phe	500	505	510	
Tyr	Lys	Phe	Thr	Gln	Lys	Gly	Gly	Ile	Leu	Lys	Leu	Gly	His	Ala	Ala	515	520	525	
Ser	Ile	Ala	Thr	Thr	Ala	Asn	Ser	Glu	Thr	Pro	Ser	Thr	Ser	Val	Gly	530	535	540	
Ser	Gln	Val	Ile	Ile	Asn	Asn	Leu	Ala	Ile	Asn	Leu	Pro	Ser	Ile	Leu	545	550	555	560
Ala	Lys	Gly	Lys	Ala	Pro	Thr	Leu	Trp	Ile	Arg	Pro	Leu	Gln	Ser	Ser	565	570	575	

Fig. 21B (con't)

Ala	Pro	Phe	Thr	Glu	Asp	Asn	Asn	Pro	Thr	Ile	Thr	Leu	Ser	Gly	Pro		
			580					585					590				
Leu	Thr	Leu	Leu	Asn	Glu	Glu	Asn	Arg	Asp	Pro	Tyr	Asp	Ser	Ile	Asp		
		595					600					605					
Leu	Ser	Glu	Pro	Leu	Gln	Asn	Ile	His	Leu	Leu	Ser	Leu	Ser	Asp	Val		
	610					615					620						
Thr	Ala	Arg	His	Ile	Asn	Thr	Asp	Asn	Phe	His	Pro	Glu	Ser	Leu	Asn		
625					630					635					640		
Ala	Thr	Glu	His	Tyr	Gly	Tyr	Gln	Gly	Ile	Trp	Ser	Pro	Tyr	Trp	Val		
				645					650						655		
Glu	Thr	Ile	Thr	Thr	Thr	Asn	Asn	Ala	Ser	Ile	Glu	Thr	Ala	Asn	Thr		
			660					665						670			
Leu	Tyr	Arg	Ala	Leu	Tyr	Ala	Asn	Trp	Thr	Pro	Leu	Gly	Tyr	Lys	Val		
		675					680					685					
Asn	Pro	Glu	Tyr	Gln	Gly	Asp	Leu	Ala	Thr	Thr	Pro	Leu	Trp	Gln	Ser		
	690					695					700						
Phe	His	Thr	Met	Phe	Ser	Leu	Leu	Arg	Ser	Tyr	Asn	Arg	Thr	Gly	Asp		
705					710					715					720		
Ser	Asp	Ile	Glu	Arg	Pro	Phe	Leu	Glu	Ile	Gln	Gly	Ile	Ala	Asp	Gly		
				725					730					735			
Leu	Phe	Val	His	Gln	Asn	Ser	Ile	Pro	Gly	Ala	Pro	Gly	Phe	Arg	Ile		
			740					745						750			
Gln	Ser	Thr	Gly	Tyr	Ser	Leu	Gln	Ala	Ser	Ser	Glu	Thr	Ser	Leu	His		
		755					760					765					
Gln	Lys	Ile	Ser	Leu	Gly	Phe	Ala	Gln	Phe	Phe	Thr	Arg	Thr	Lys	Glu		
	770					775					780						
Ile	Gly	Ser	Ser	Asn	Asn	Val	Ser	Ala	His	Asn	Thr	Val	Ser	Ser	Leu		
785					790					795					800		
Tyr	Val	Glu	Leu	Pro	Trp	Phe	Gln	Glu	Ala	Phe	Ala	Thr	Ser	His	Ser		
				805					810						815		
Leu	Ala	Tyr	Gly	Tyr	Gly	Asp	His	His	Leu	His	Ala	Tyr	Ile	Arg	His		
			820					825						830			
Ile	Lys	Asn	Arg	Ala	Glu	Gly	Thr	Cys	Tyr	Ser	His	Thr	Leu	Ala	Ala		
		835					840						845				
Ala	Ile	Gly	Cys	Ser	Phe	Pro	Trp	Gln	Gln	Lys	Ser	Tyr	Leu	His	Leu		
	850						855					860					

Fig. 21B (con't)

Ser	Pro	Phe	Val	Gln	Ala	Ile	Ala	Ile	Arg	Ser	His	Gln	Thr	Ala	Phe	
865					870					875					880	
Glu	Glu	Ile	Gly	Asp	Asn	Pro	Arg	Lys	Phe	Val	Ser	Gln	Lys	Pro	Phe	
				885					890					895		
Tyr	Asn	Leu	Thr	Leu	Pro	Leu	Gly	Ile	Gln	Gly	Lys	Trp	Gln	Ser	Lys	
			900					905					910			
Phe	His	Val	Pro	Thr	Glu	Trp	Thr	Leu	Glu	Leu	Ser	Tyr	Gln	Pro	Val	
		915					920					925				
Leu	Tyr	Gln	Gln	Asn	Pro	Gln	Ile	Gly	Val	Thr	Leu	Leu	Ala	Ser	Gly	
	930					935					940					
Gly	Ser	Trp	Asp	Ile	Leu	Gly	His	Asn	Tyr	Val	Arg	Asn	Ala	Leu	Gly	
945					950					955					960	
Tyr	Lys	Val	His	Asn	Gln	Thr	Ala	Leu	Phe	Arg	Ser	Leu	Asp	Leu	Phe	
				965					970					975		
Leu	Asp	Tyr	Gln	Gly	Ser	Val	Ser	Ser	Ser	Thr	Ser	Thr	His	His	Leu	
			980					985					990			
Gln	Ala	Gly	Ser	Thr	Leu	Lys	Phe									
		995					1000									

Figure 22 (RY-45)

Restriction enzyme analysis of CPN100626

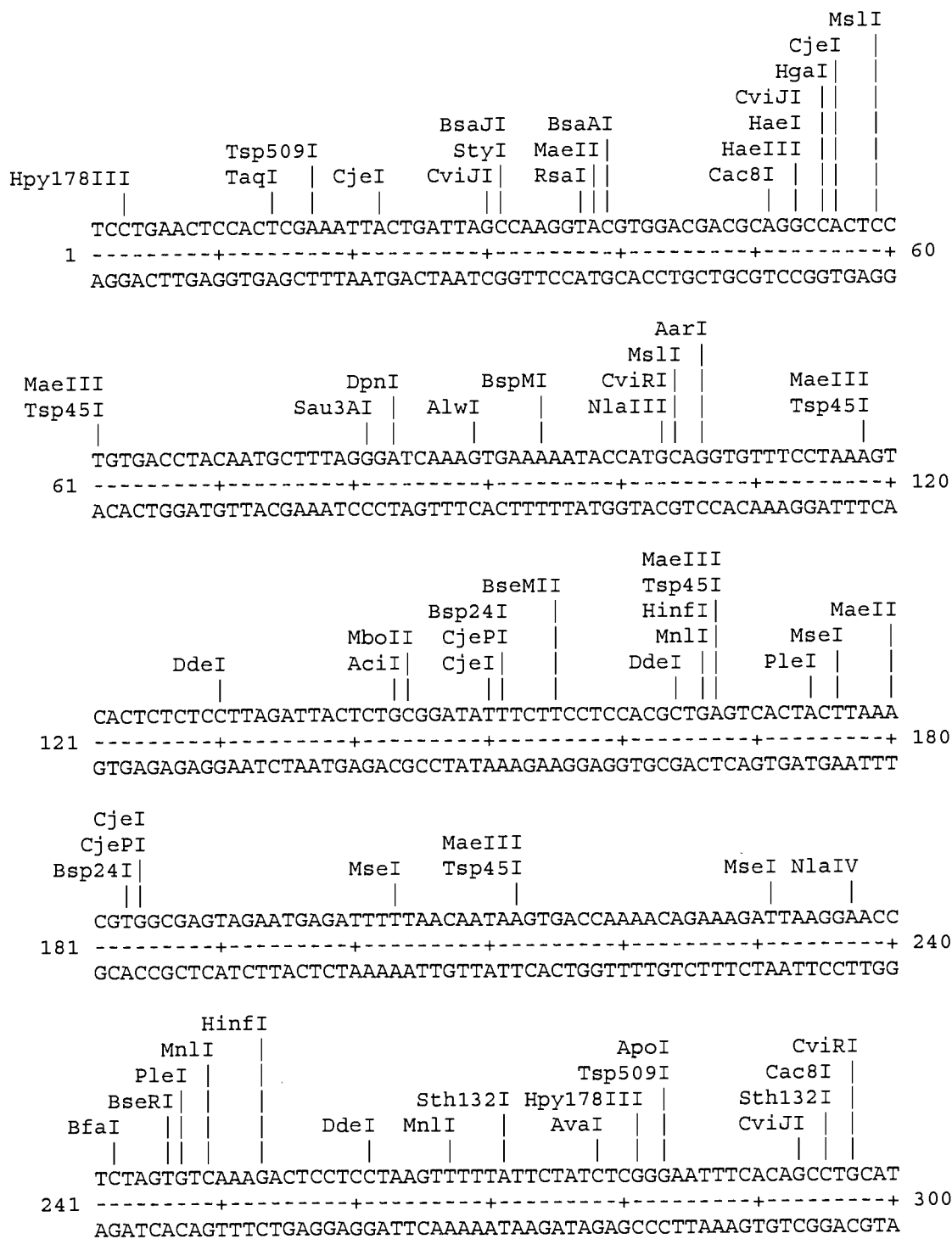


Fig. 22 (con't)

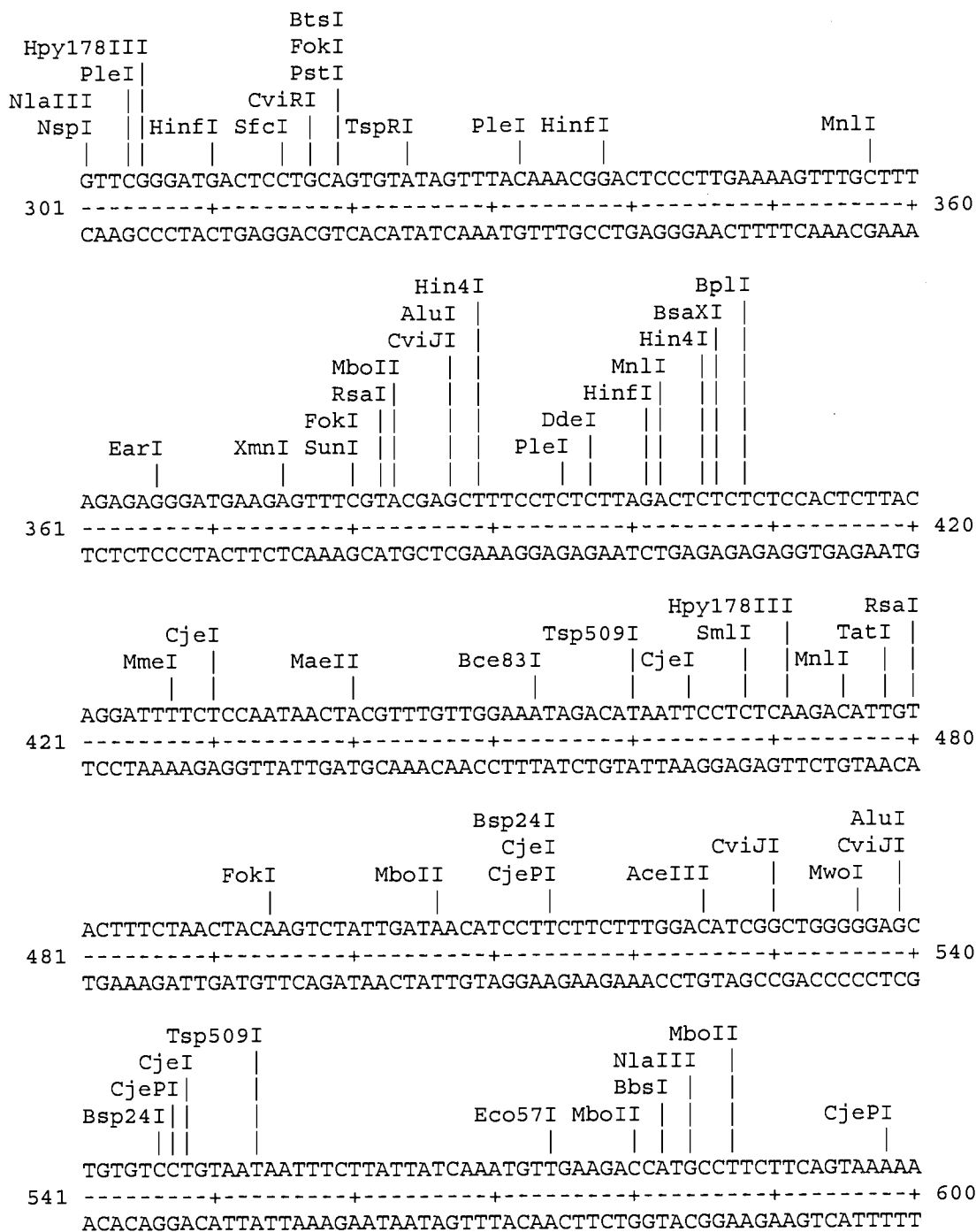




Fig. 22 (con't)

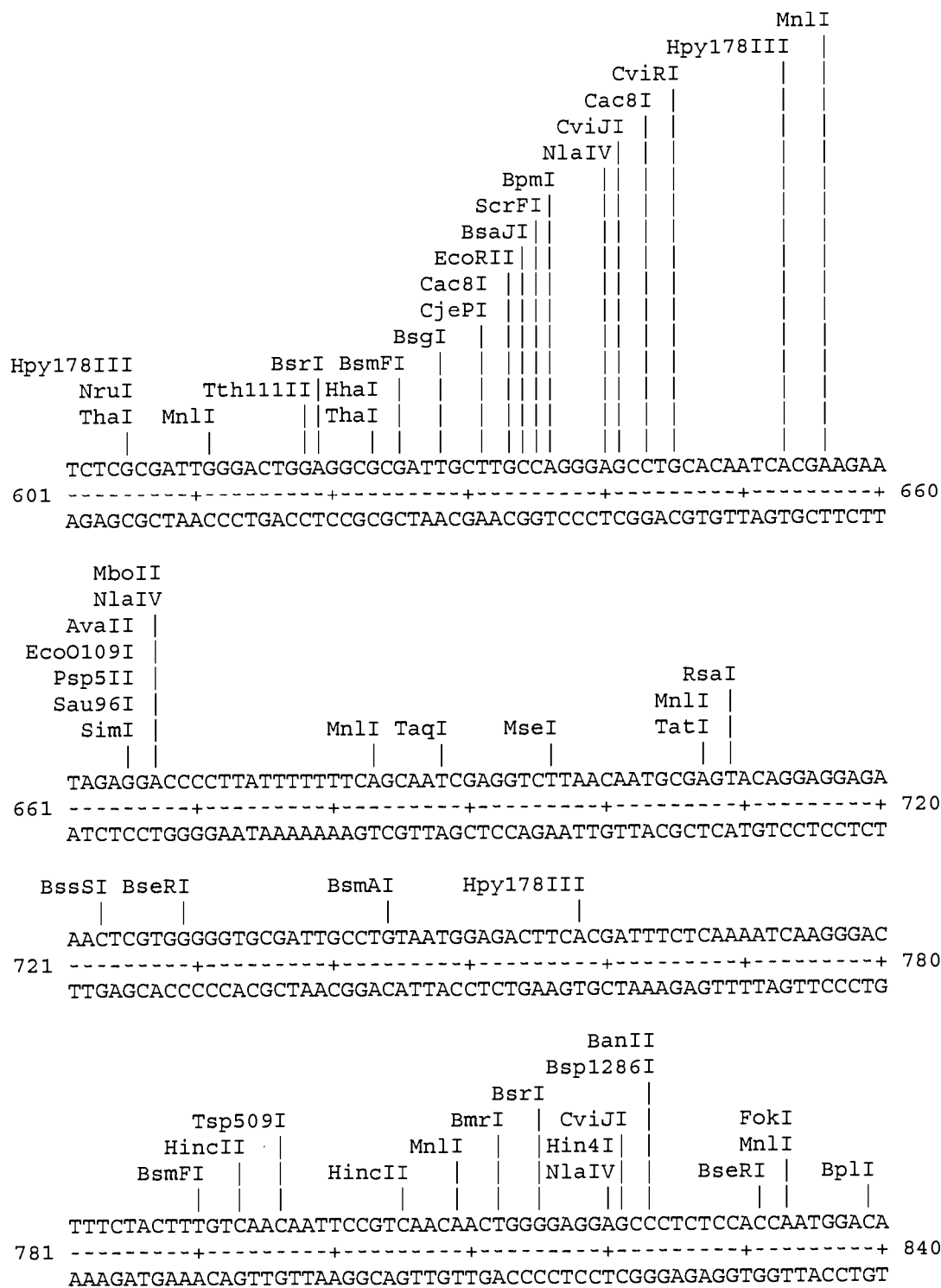


Fig. 22 (con't)

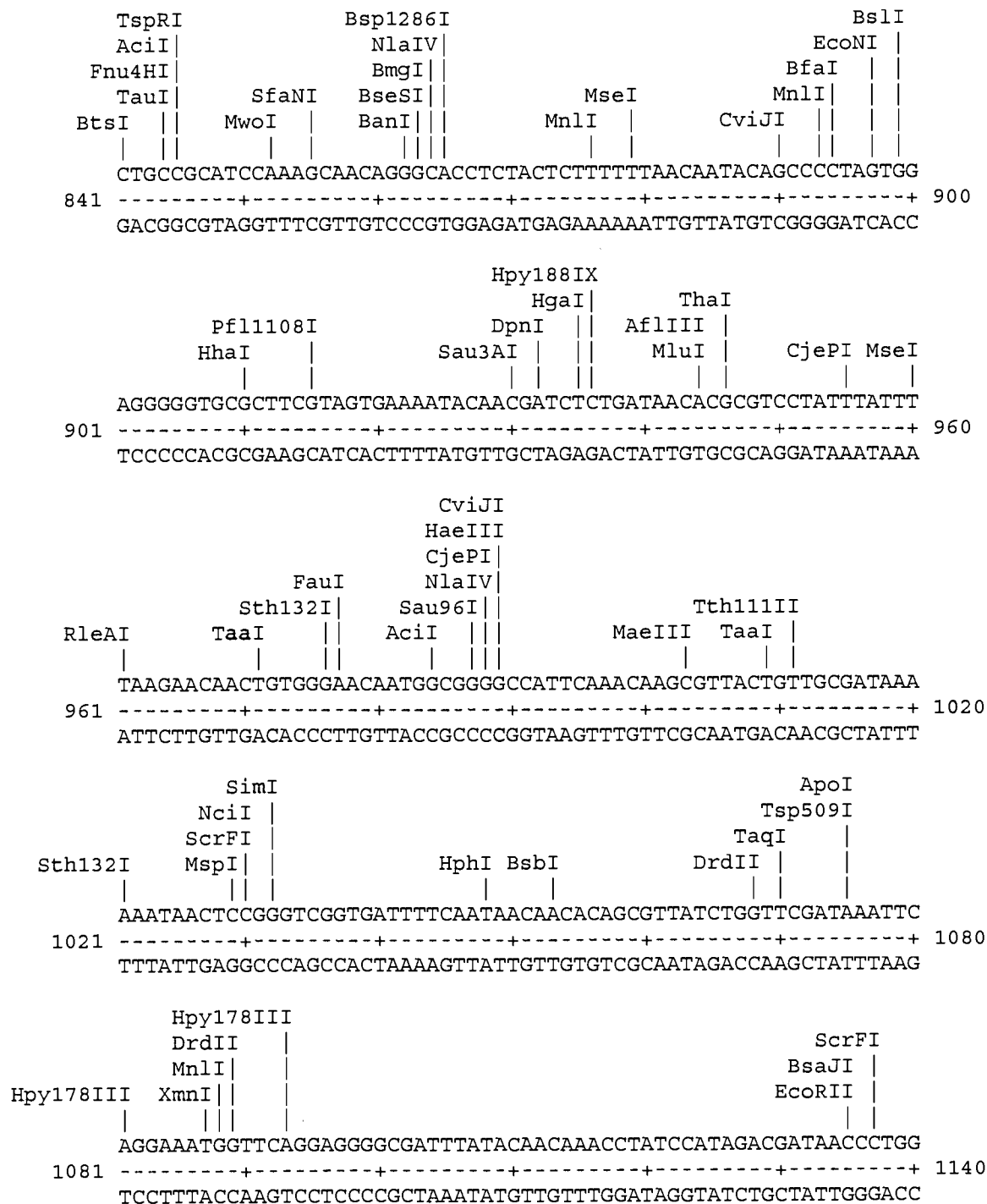


Fig. 22 (con't)

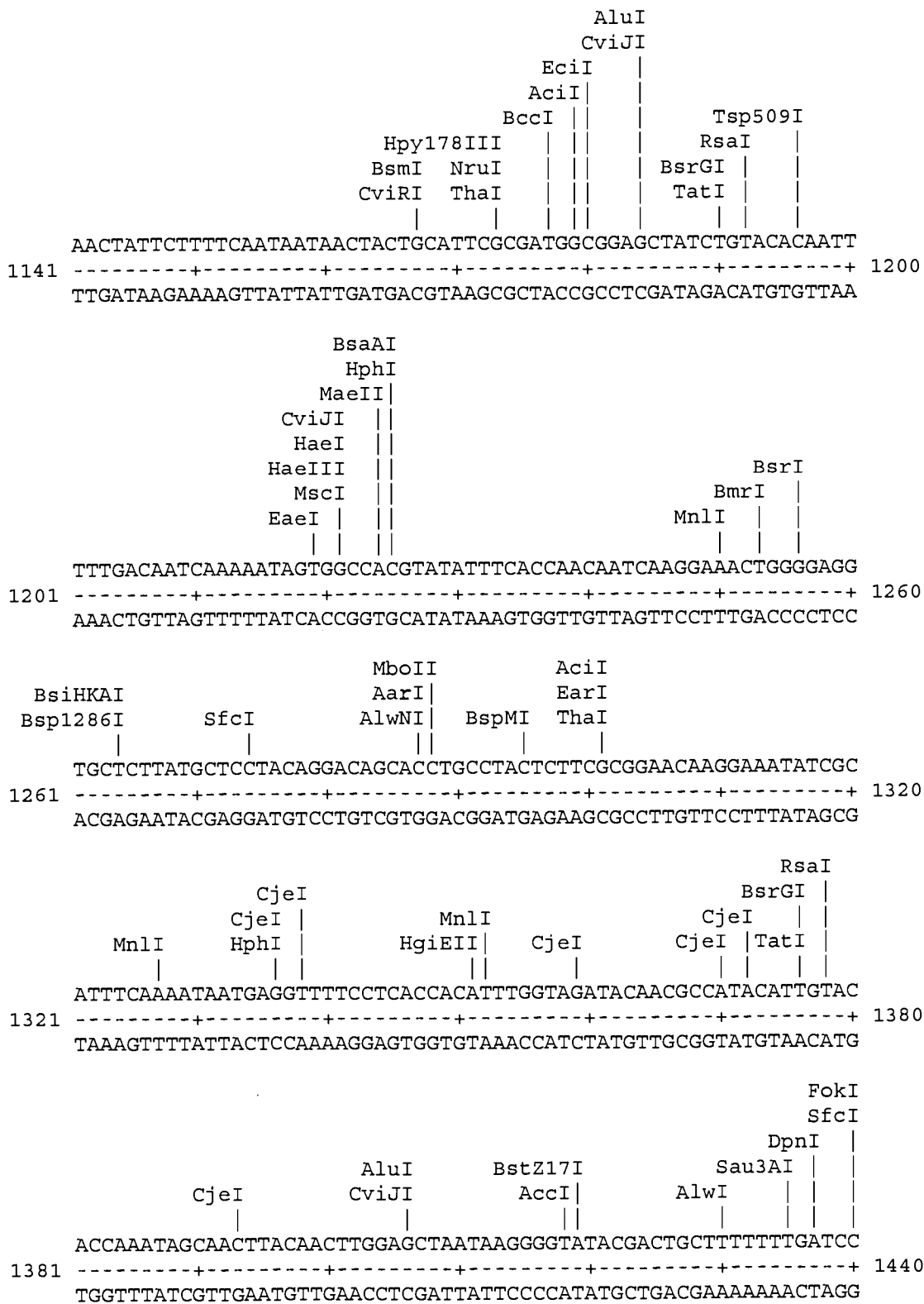


Fig. 22 (con't)

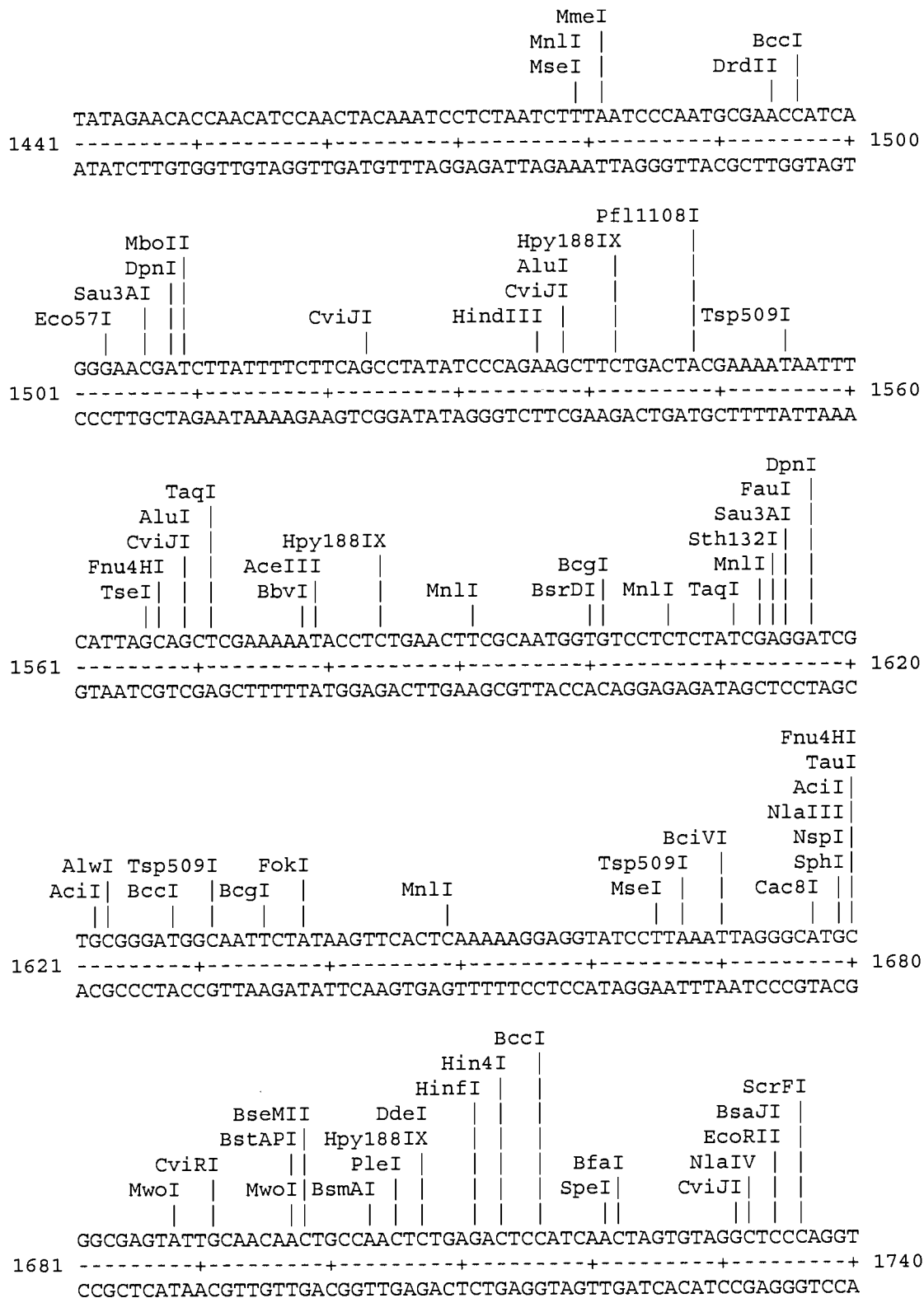


Fig. 22 (con't)

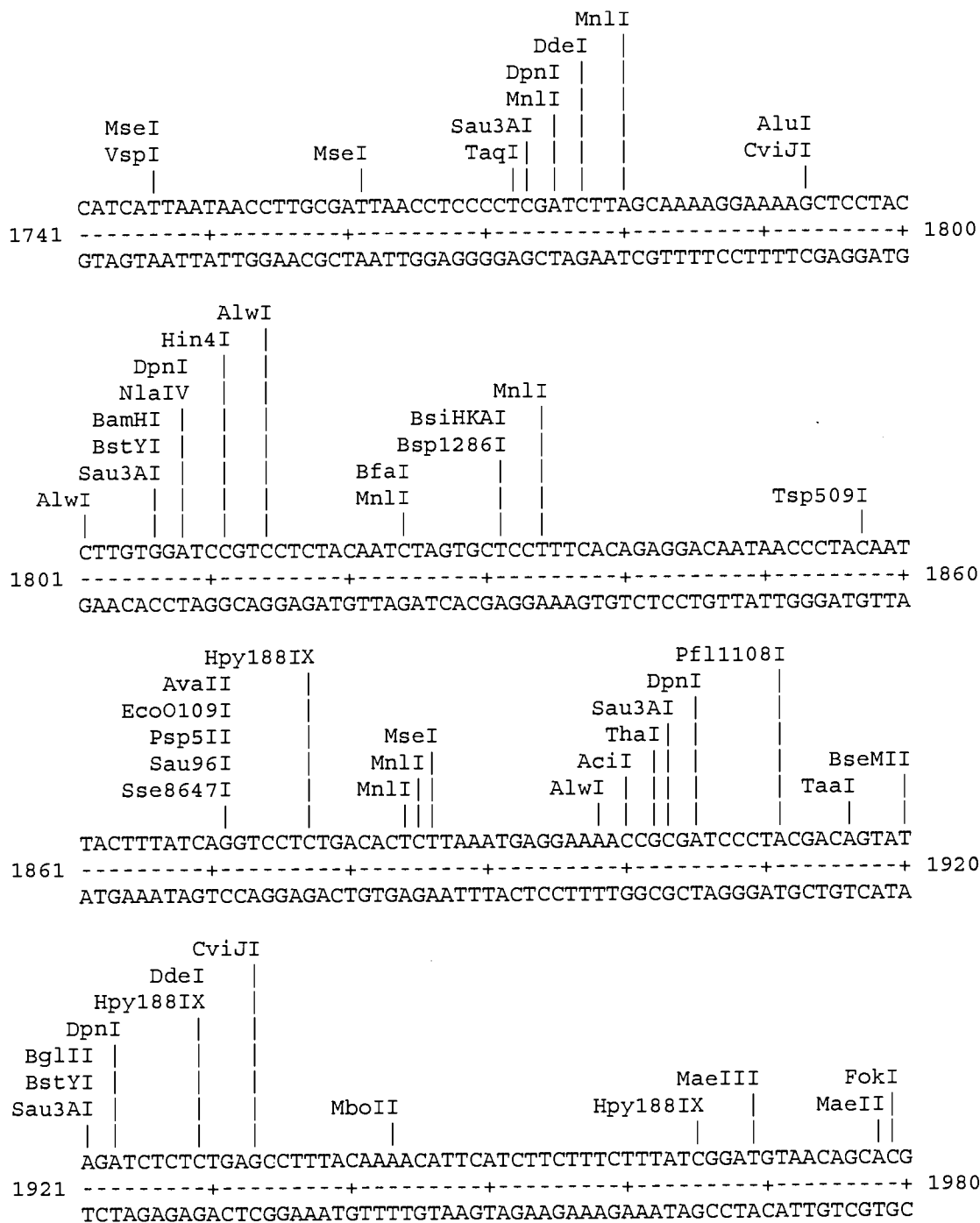


Fig. 22 (con't)

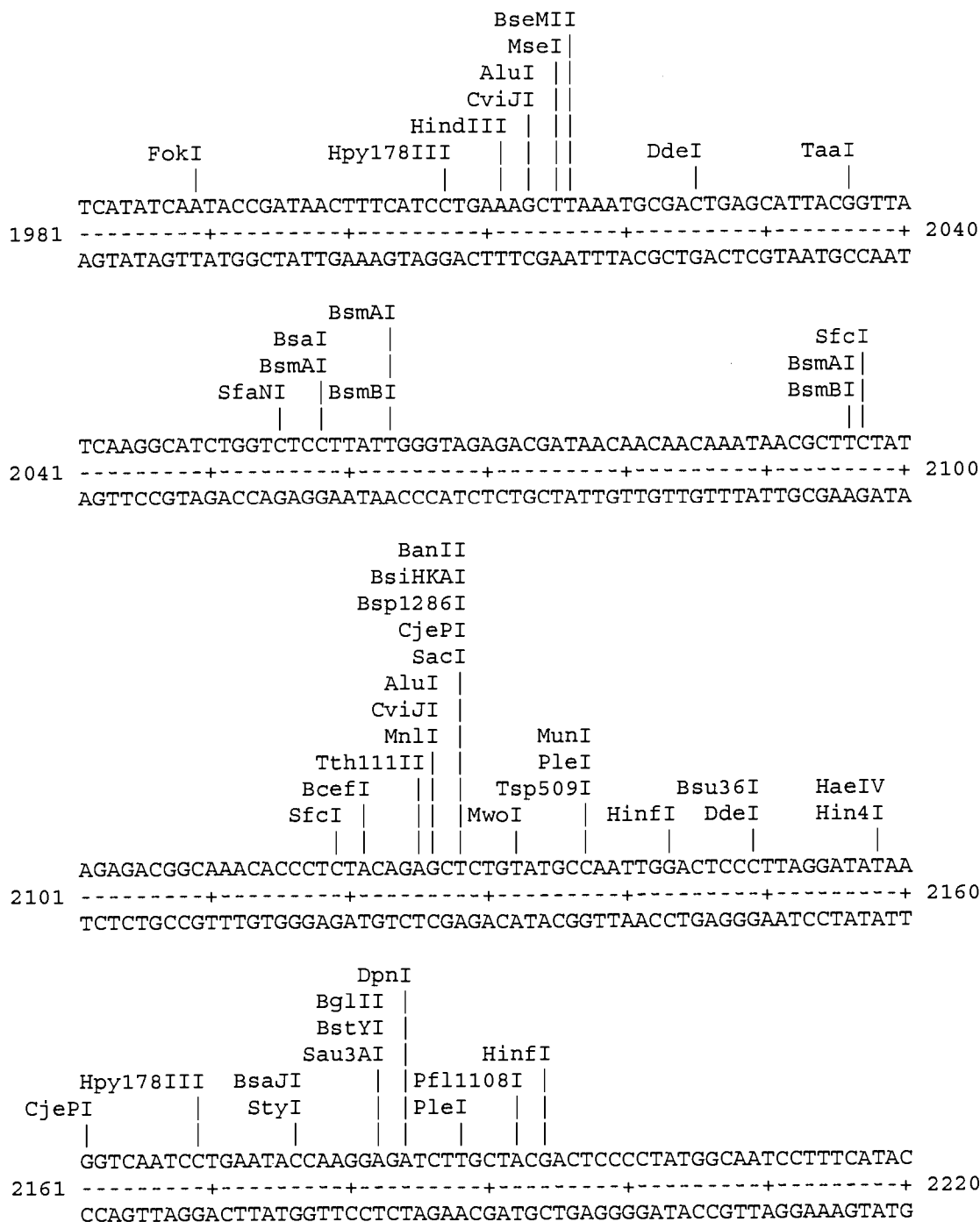


Fig. 22 (con't)

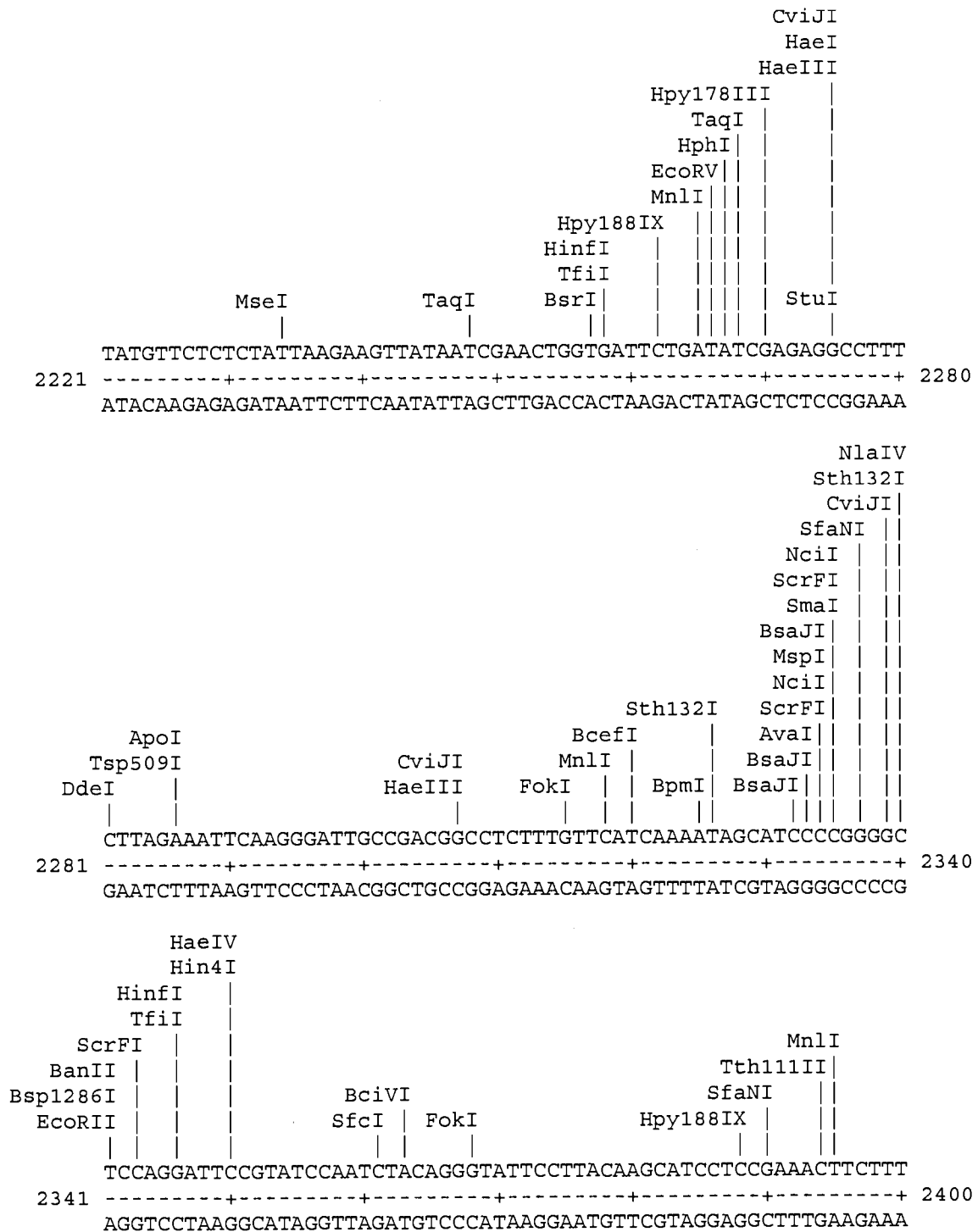


Fig. 22 (con't)

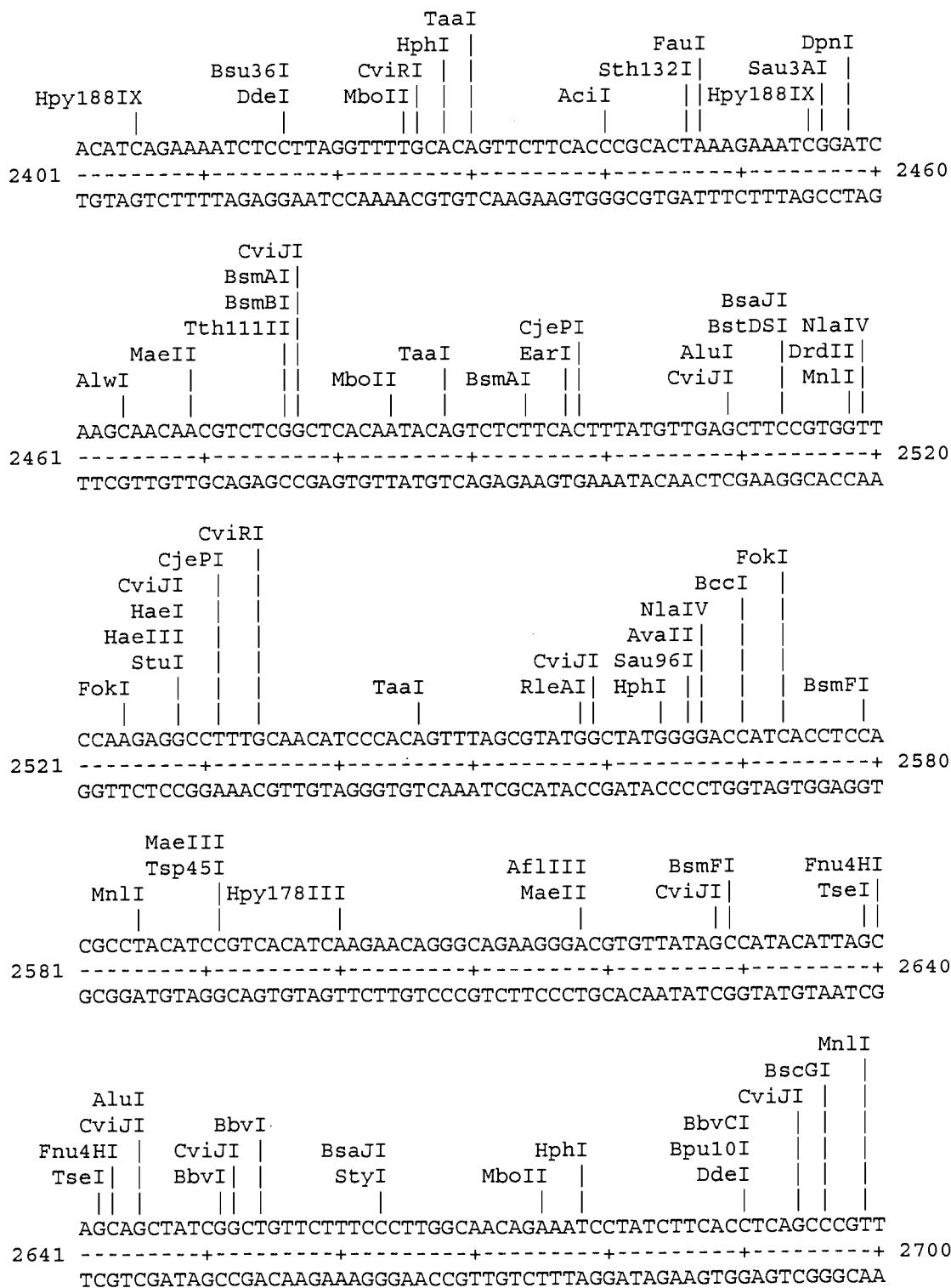




Fig. 22 (con't)

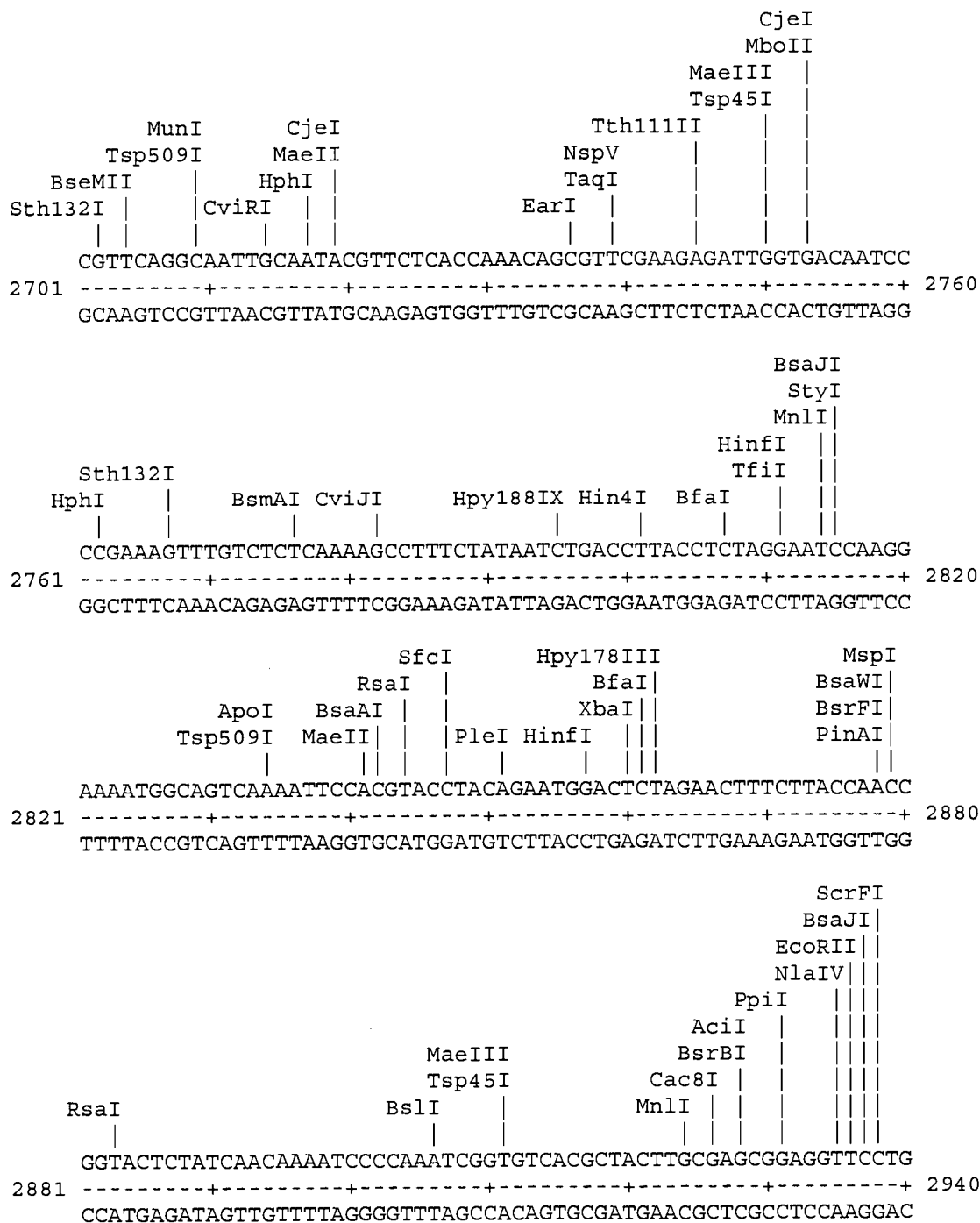


Fig. 22 (con't)

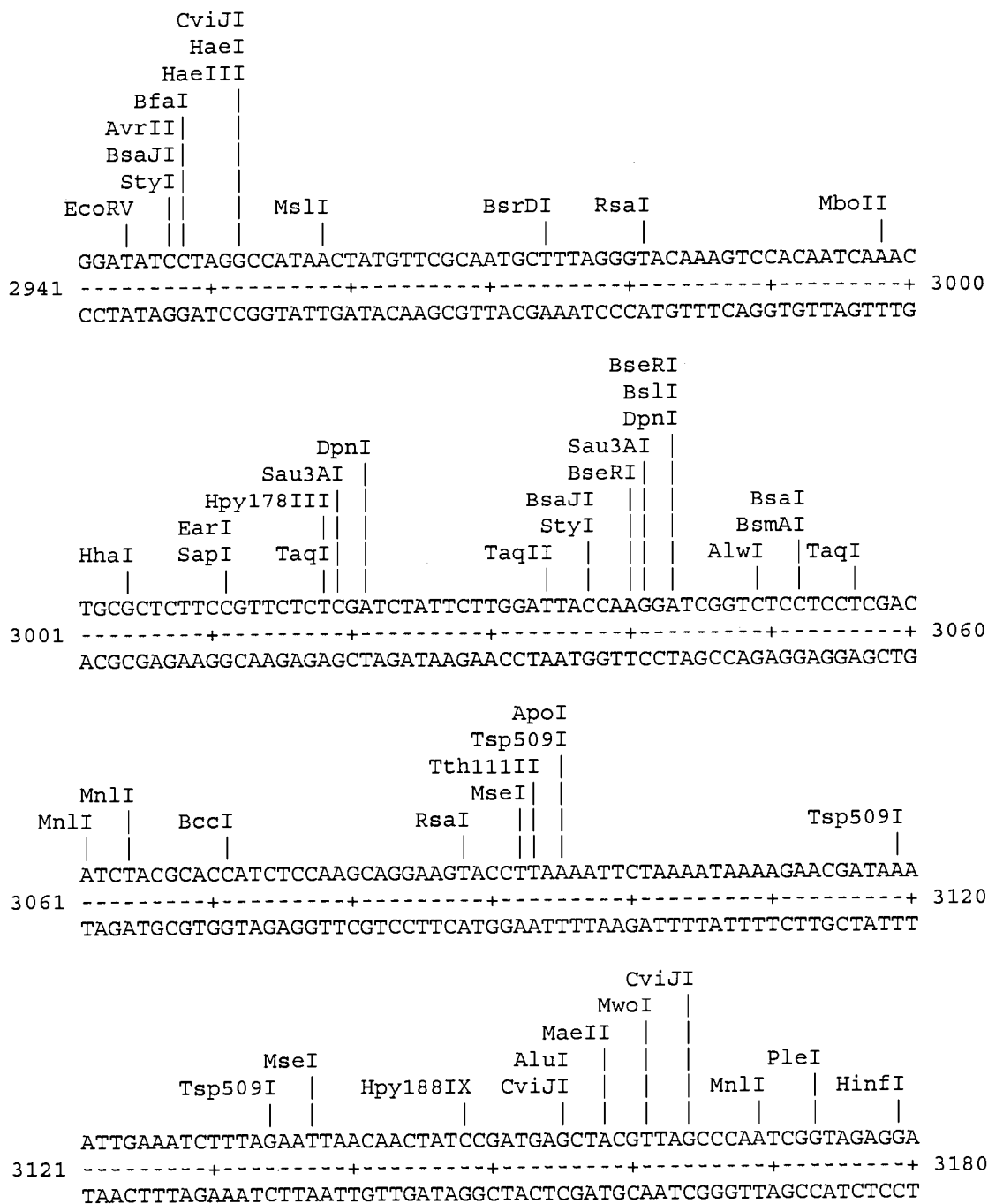


Fig. 22 (con't)

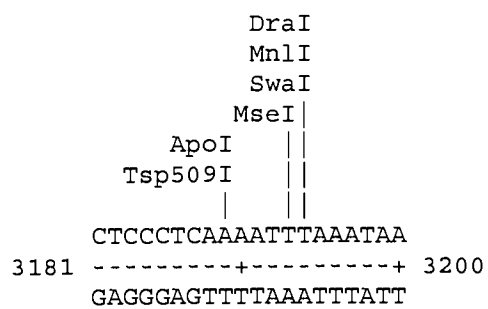


Figure 23:

```

tagacactat aaaacaaatt atagacaaaa aatctagcat tgatttattc agaatatattc 60

tttctatttg tgaacgagta tgcgcttttt ttgcttcgga atg ttg ctt cct ttt 115
                                         Met Leu Leu Pro Phe
                                         1           5

act ttt gta ttg gct aat gaa ggt ctc caa ctt cct ttg gag acc tat 163
Thr Phe Val Leu Ala Asn Glu Gly Leu Gln Leu Pro Leu Glu Thr Tyr
                        10                15                20

att aca tta agt cct gaa tat caa gca gcc cct caa gta ggg ttt act 211
Ile Thr Leu Ser Pro Glu Tyr Gln Ala Ala Pro Gln Val Gly Phe Thr
                        25                30                35

cat aac caa aat caa gat ctc gca att gtc ggg aat cac aat gat ttc 259
His Asn Gln Asn Gln Asp Leu Ala Ile Val Gly Asn His Asn Asp Phe
                        40                45                50

atc ttg gac tat aag tac tat cgg tcg aat gga ggt gct ctt acc tgt 307
Ile Leu Asp Tyr Lys Tyr Tyr Arg Ser Asn Gly Gly Ala Leu Thr Cys
                        55                60                65

aag aat ctt ctg atc tct gaa aat ata ggg aat gtc ttc ttt gag aag 355
Lys Asn Leu Leu Ile Ser Glu Asn Ile Gly Asn Val Phe Phe Glu Lys
                        70                75                80                85

aat gtc tgt ccc aat tct ggc ggg gca att tat gct gct caa aat tgc 403
Asn Val Cys Pro Asn Ser Gly Gly Ala Ile Tyr Ala Ala Gln Asn Cys
                        90                95                100

acg atc tcc aag aat cag aac tat gca ttt act aca aac ttg gtc tct 451
Thr Ile Ser Lys Asn Gln Asn Tyr Ala Phe Thr Thr Asn Leu Val Ser
                        105                110                115

gac aat cct aca gcc act gcg gga tca cta ttg ggt gga gct ctc ttt 499
Asp Asn Pro Thr Ala Thr Ala Gly Ser Leu Leu Gly Gly Ala Leu Phe
                        120                125                130

gcc ata aat tgc tct att act aat aac cta gga cag gga act ttc gtt 547
Ala Ile Asn Cys Ser Ile Thr Asn Asn Leu Gly Gln Gly Thr Phe Val
                        135                140                145

gac aat ctc gct tta aat aag ggg ggt gcc ctc tat act gag acg aac 595
Asp Asn Leu Ala Leu Asn Lys Gly Gly Ala Leu Tyr Thr Glu Thr Asn
                        150                155                160                165

tta tct att aaa gac aat aaa ggc ccg atc ata atc aag cag aat cgg 643
Leu Ser Ile Lys Asp Asn Lys Gly Pro Ile Ile Ile Lys Gln Asn Arg
                        170                175                180

gca cta aat tcg gac agt tta gga gga ggg att tat agt ggg aac tct 691
Ala Leu Asn Ser Asp Ser Leu Gly Gly Gly Ile Tyr Ser Gly Asn Ser
                        185                190                195

```

Fig. 23 (con't)

cta aat ata gag gga aat tct gga gct ata cag atc aca agc aac tct	739
Leu Asn Ile Glu Gly Asn Ser Gly Ala Ile Gln Ile Thr Ser Asn Ser	
200 205 210	
tca gga tct ggg gga ggc ata ttt tct acc caa aca ctc acg atc tcc	787
Ser Gly Ser Gly Gly Ile Phe Ser Thr Gln Thr Leu Thr Ile Ser	
215 220 225	
tcg aat aaa aaa ctc ata gaa atc agt gaa aat tcc gcg ttc gca aat	835
Ser Asn Lys Lys Leu Ile Glu Ile Ser Glu Asn Ser Ala Phe Ala Asn	
230 235 240 245	
aac tat gga tcg aac ttc aat cca gga gga gga ggt ctt act acc acc	883
Asn Tyr Gly Ser Asn Phe Asn Pro Gly Gly Gly Gly Leu Thr Thr Thr	
250 255 260	
ttt tgc acg ata ttg aac aac cga gaa ggg gta ctc ttt aac aat aac	931
Phe Cys Thr Ile Leu Asn Asn Arg Glu Gly Val Leu Phe Asn Asn Asn	
265 270 275	
caa agc cag agc aac ggt gga gcc att cat gcg aaa tct atc att atc	979
Gln Ser Gln Ser Asn Gly Gly Ala Ile His Ala Lys Ser Ile Ile Ile	
280 285 290	
aaa gaa aat ggt cct gta tac ttt tta aat aac act gca act cgg gga	1027
Lys Glu Asn Gly Pro Val Tyr Phe Leu Asn Asn Thr Ala Thr Arg Gly	
295 300 305	
ggg gct ctc ctc aac tta tca gca ggt tct gga aac gga agc ttc atc	1075
Gly Ala Leu Leu Asn Leu Ser Ala Gly Ser Gly Asn Gly Ser Phe Ile	
310 315 320 325	
tta tct gca gat aat gga gat att atc ttt aac aat aat acg gcc tcc	1123
Leu Ser Ala Asp Asn Gly Asp Ile Ile Phe Asn Asn Asn Thr Ala Ser	
330 335 340	
aag cat gcc ctc aat cct cca tac aga aac gcc att cac tcg act cct	1171
Lys His Ala Leu Asn Pro Pro Tyr Arg Asn Ala Ile His Ser Thr Pro	
345 350 355	
aat atg aat ctg caa ata gga gcc cgt ccc ggc tat cga gtg ctg ttc	1219
Asn Met Asn Leu Gln Ile Gly Ala Arg Pro Gly Tyr Arg Val Leu Phe	
360 365 370	
tat gat ccc ata gaa cat gag ctc cct tcc tcc ttc ccc ata ctc ttt	1267
Tyr Asp Pro Ile Glu His Glu Leu Pro Ser Ser Phe Pro Ile Leu Phe	
375 380 385	
aat ttc gaa acc ggt cat aca ggt aca gtt tta ttt tca ggg gaa cat	1315
Asn Phe Glu Thr Gly His Thr Gly Thr Val Leu Phe Ser Gly Glu His	
390 395 400 405	

Fig. 23 (con't)

gta cac cag aac ttt acc gat gaa atg aat ttc ttt tcc tat tta agg	1363
Val His Gln Asn Phe Thr Asp Glu Met Asn Phe Phe Ser Tyr Leu Arg	
410 415 420	
aac act tcg gaa cta cgt caa gga gtc ctt gct gtt gaa gat ggt gcg	1411
Asn Thr Ser Glu Leu Arg Gln Gly Val Leu Ala Val Glu Asp Gly Ala	
425 430 435	
ggg ctg gcc tgc tat aag ttc ttc caa cga gga ggc act cta ctt cta	1459
Gly Leu Ala Cys Tyr Lys Phe Phe Gln Arg Gly Gly Thr Leu Leu Leu	
440 445 450	
ggt caa ggt gcg gtg atc acg aca gca gga acg att ccc aca cca tcc	1507
Gly Gln Gly Ala Val Ile Thr Thr Ala Gly Thr Ile Pro Thr Pro Ser	
455 460 465	
tca aca cca acg aca gta gga agt act ata act tta aat cac att gcc	1555
Ser Thr Pro Thr Thr Val Gly Ser Thr Ile Thr Leu Asn His Ile Ala	
470 475 480 485	
att gac ctt cct tct att ctt tct ttt caa gct cag gct cca aaa att	1603
Ile Asp Leu Pro Ser Ile Leu Ser Phe Gln Ala Gln Ala Pro Lys Ile	
490 495 500	
tgg att tac ccc aca aaa aca gga tct acc tat act gaa gat tcc aac	1651
Trp Ile Tyr Pro Thr Lys Thr Gly Ser Thr Tyr Thr Glu Asp Ser Asn	
505 510 515	
ccg aca atc aca atc tca gga act ctc acc tta cgc aac agc aac aac	1699
Pro Thr Ile Thr Ile Ser Gly Thr Leu Thr Leu Arg Asn Ser Asn Asn	
520 525 530	
gaa gat ccc tac gat agt ctg gat ctc tcg cac tct ctt gag aaa gtt	1747
Glu Asp Pro Tyr Asp Ser Leu Asp Leu Ser His Ser Leu Glu Lys Val	
535 540 545	
ccc ctt ctt tat att gtc gat gtc gct gca caa aaa att aac tct tcg	1795
Pro Leu Leu Tyr Ile Val Asp Val Ala Ala Gln Lys Ile Asn Ser Ser	
550 555 560 565	
caa ctg gat cta tcc aca tta aat tct ggc gaa cac tat ggg tat caa	1843
Gln Leu Asp Leu Ser Thr Leu Asn Ser Gly Glu His Tyr Gly Tyr Gln	
570 575 580	
ggc atc tgg tcg acc tat tgg gta gaa act aca aca atc acg aac cct	1891
Gly Ile Trp Ser Thr Tyr Trp Val Glu Thr Thr Thr Ile Thr Asn Pro	
585 590 595	
aca tct cta cta ggc gcg aat aca aaa cac aag ctg ctc tat gca aac	1939
Thr Ser Leu Leu Gly Ala Asn Thr Lys His Lys Leu Leu Tyr Ala Asn	
600 605 610	

Fig. 23 (con't)

tgg tct cct cta ggc tac cgt cct cat ccc gaa cgt cga gga gaa ttc	1987
Trp Ser Pro Leu Gly Tyr Arg Pro His Pro Glu Arg Arg Gly Glu Phe	
615 620 625	
att acg aat gcc ttg tgg caa tcg gca tat acg gct ctt gca gga ctc	2035
Ile Thr Asn Ala Leu Trp Gln Ser Ala Tyr Thr Ala Leu Ala Gly Leu	
630 635 640 645	
cac tcc ctc tcc tcc tgg gat gaa gag aag ggt cat gca gct tcc cta	2083
His Ser Leu Ser Ser Trp Asp Glu Glu Lys Gly His Ala Ala Ser Leu	
650 655 660	
caa ggc att ggt ctt ctg gtt cat caa aaa gac aaa aac ggt ttt aag	2131
Gln Gly Ile Gly Leu Leu Val His Gln Lys Asp Lys Asn Gly Phe Lys	
665 670 675	
gga ttt cgt agt cat atg aca ggt tat agt gct acc acc gaa gca acc	2179
Gly Phe Arg Ser His Met Thr Gly Tyr Ser Ala Thr Thr Glu Ala Thr	
680 685 690	
tct tct caa agt ccg aat ttc tct tta gga ttt gct cag ttc ttc tcc	2227
Ser Ser Gln Ser Pro Asn Phe Ser Leu Gly Phe Ala Gln Phe Phe Ser	
695 700 705	
aaa gct aaa gaa cat gaa tct caa aat agc acg tcc tct cac cac tat	2275
Lys Ala Lys Glu His Glu Ser Gln Asn Ser Thr Ser Ser His His Tyr	
710 715 720 725	
ttc tct gga atg tgc ata gca aaa tac tct ctt caa aga gtg ata cgt	2323
Phe Ser Gly Met Cys Ile Ala Lys Tyr Ser Leu Gln Arg Val Ile Arg	
730 735 740	
cta tct gtg tct ctt gct tat atg ttt acc tcg gaa cat acc cat aca	2371
Leu Ser Val Ser Leu Ala Tyr Met Phe Thr Ser Glu His Thr His Thr	
745 750 755	
atg tat cag ggt ctc ctg gaa ggg aac tct cag gga tct ttc cac aac	2419
Met Tyr Gln Gly Leu Leu Glu Gly Asn Ser Gln Gly Ser Phe His Asn	
760 765 770	
cat acc tta gca ggg gct ctc tcc tgt gtt ttc tta cct caa cct cac	2467
His Thr Leu Ala Gly Ala Leu Ser Cys Val Phe Leu Pro Gln Pro His	
775 780 785	
ggc gag tcc ctg cag atc tat ccc ttt att act gcc tta gcc atc cga	2515
Gly Glu Ser Leu Gln Ile Tyr Pro Phe Ile Thr Ala Leu Ala Ile Arg	
790 795 800 805	
gga aat ctt gct gcg ttt caa gaa tct gga gac cat gct cgg gaa ttt	2563
Gly Asn Leu Ala Ala Phe Gln Glu Ser Gly Asp His Ala Arg Glu Phe	
810 815 820	

Fig. 23 (con't)

tcc cta cac cgc ccc cta acg gac gtc tcc ctc cct gta gga atc cgc	2611
Ser Leu His Arg Pro Leu Thr Asp Val Ser Leu Pro Val Gly Ile Arg	
825 830 835	
 gct tct tgg aag aac cac cac cga gtt ccc cta gtc tgg ctc aca gaa	2659
Ala Ser Trp Lys Asn His His Arg Val Pro Leu Val Trp Leu Thr Glu	
840 845 850	
 att tcc tat cgc tct act ctc tat agg caa gat cct gaa ctc cac tcg	2707
Ile Ser Tyr Arg Ser Thr Leu Tyr Arg Gln Asp Pro Glu Leu His Ser	
855 860 865	
 aaa tta ctg att agc caa ggt acg tgg acg acg cag gcc act cct gtg	2755
Lys Leu Leu Ile Ser Gln Gly Thr Trp Thr Thr Gln Ala Thr Pro Val	
870 875 880 885	
 acc tac aat gct tta ggg atc aaa gtg aaa aat acc atg cag gtg ttt	2803
Thr Tyr Asn Ala Leu Gly Ile Lys Val Lys Asn Thr Met Gln Val Phe	
890 895 900	
 cct aaa gtc act ctc tcc tta gat tac tct gcg gat att tct tcc tcc	2851
Pro Lys Val Thr Leu Ser Leu Asp Tyr Ser Ala Asp Ile Ser Ser Ser	
905 910 915	
 acg ctg agt cac tac tta aac gtg gcg agt aga atg aga ttt	2893
Thr Leu Ser His Tyr Leu Asn Val Ala Ser Arg Met Arg Phe	
920 925 930	
 taacaataag tgaccaaaac agaaagatta aggaacctct agtgtcaaag actcctccta	2953
agttttttatt ctatctcgga aatttcacag cctgcatgtt cgggatg	3000



Figure 24 (RY-46)

## Restriction enzyme analysis of CPN100628

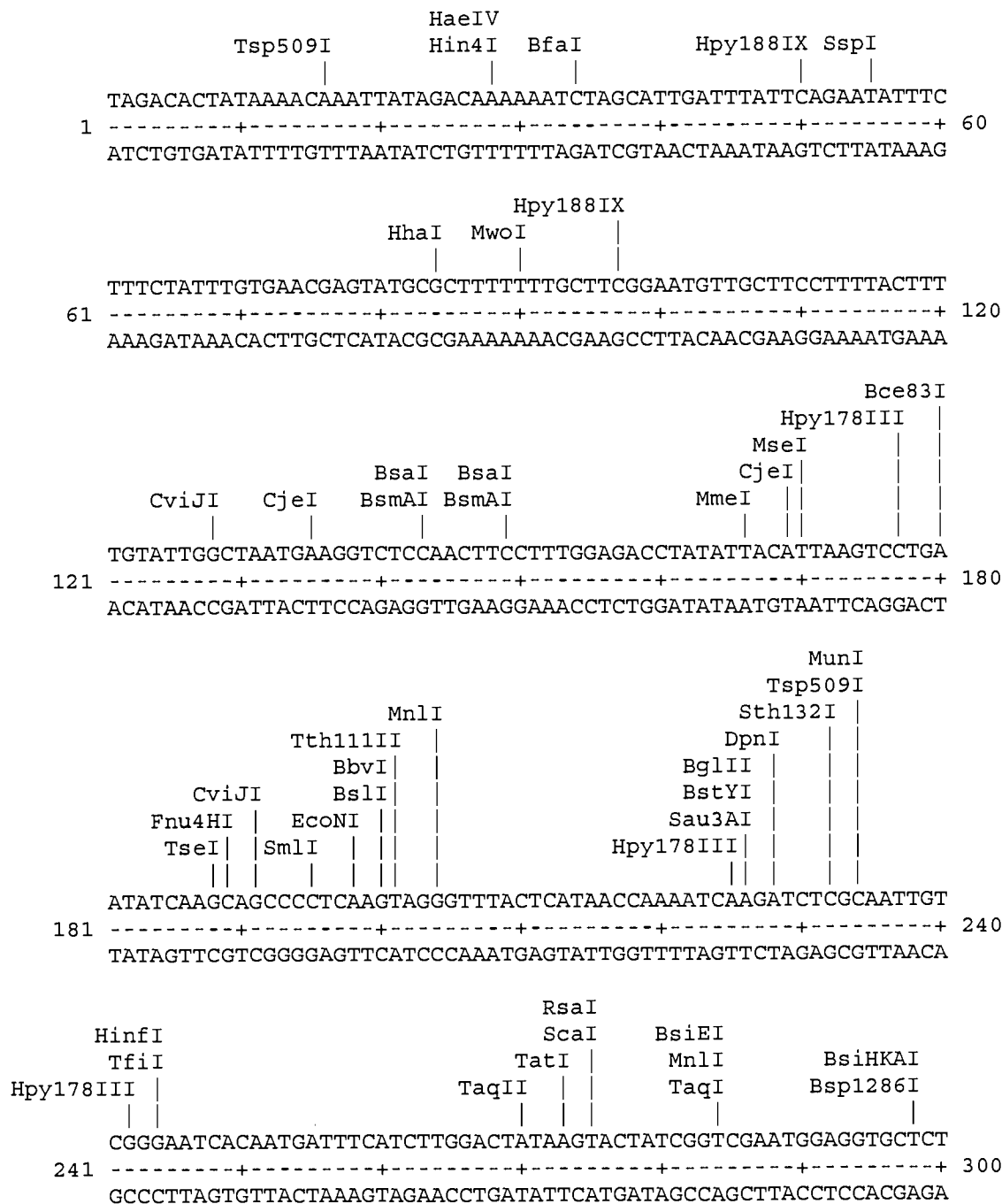


Fig. 24 (con't)

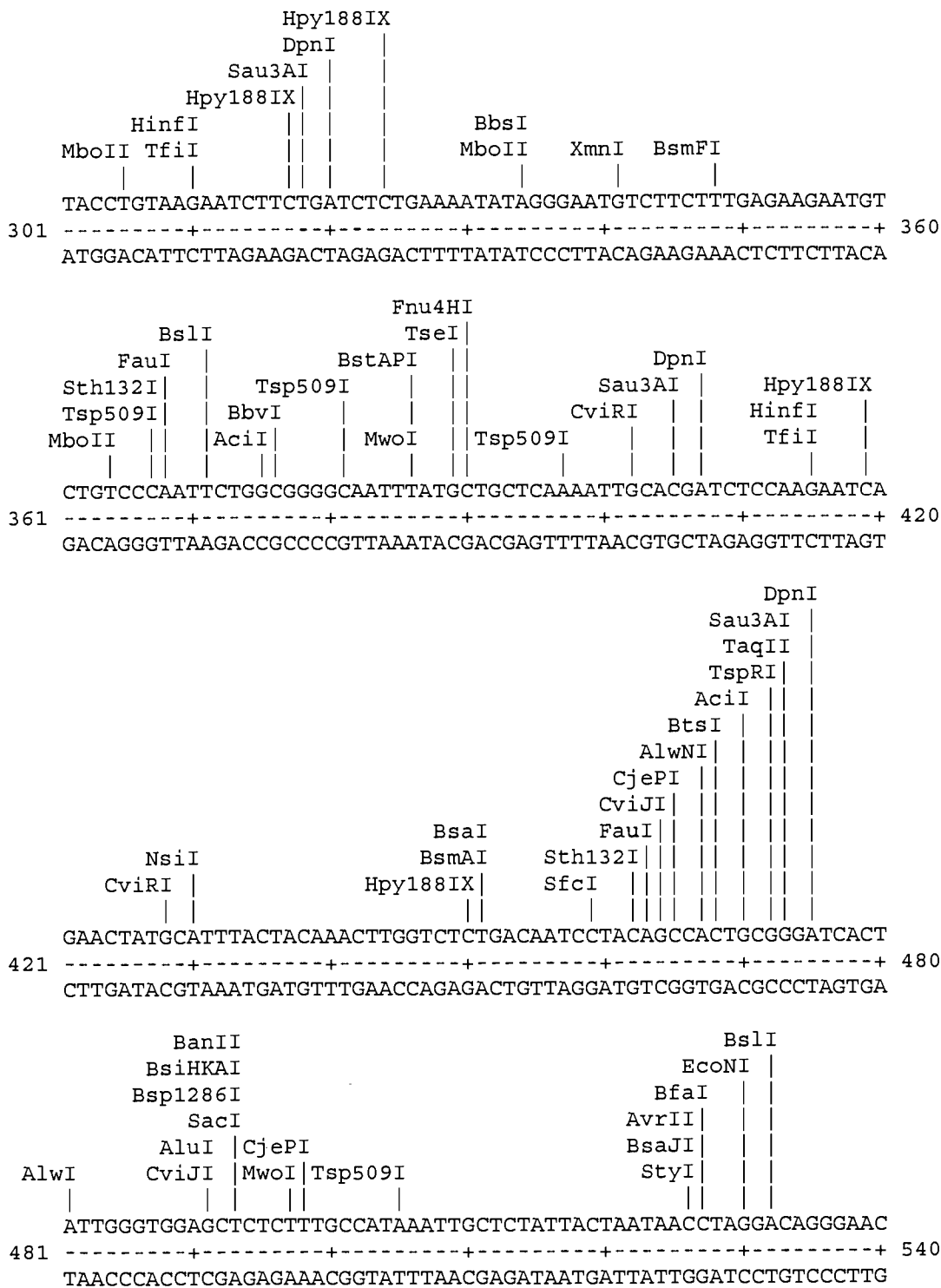


Fig. 24 (con't)

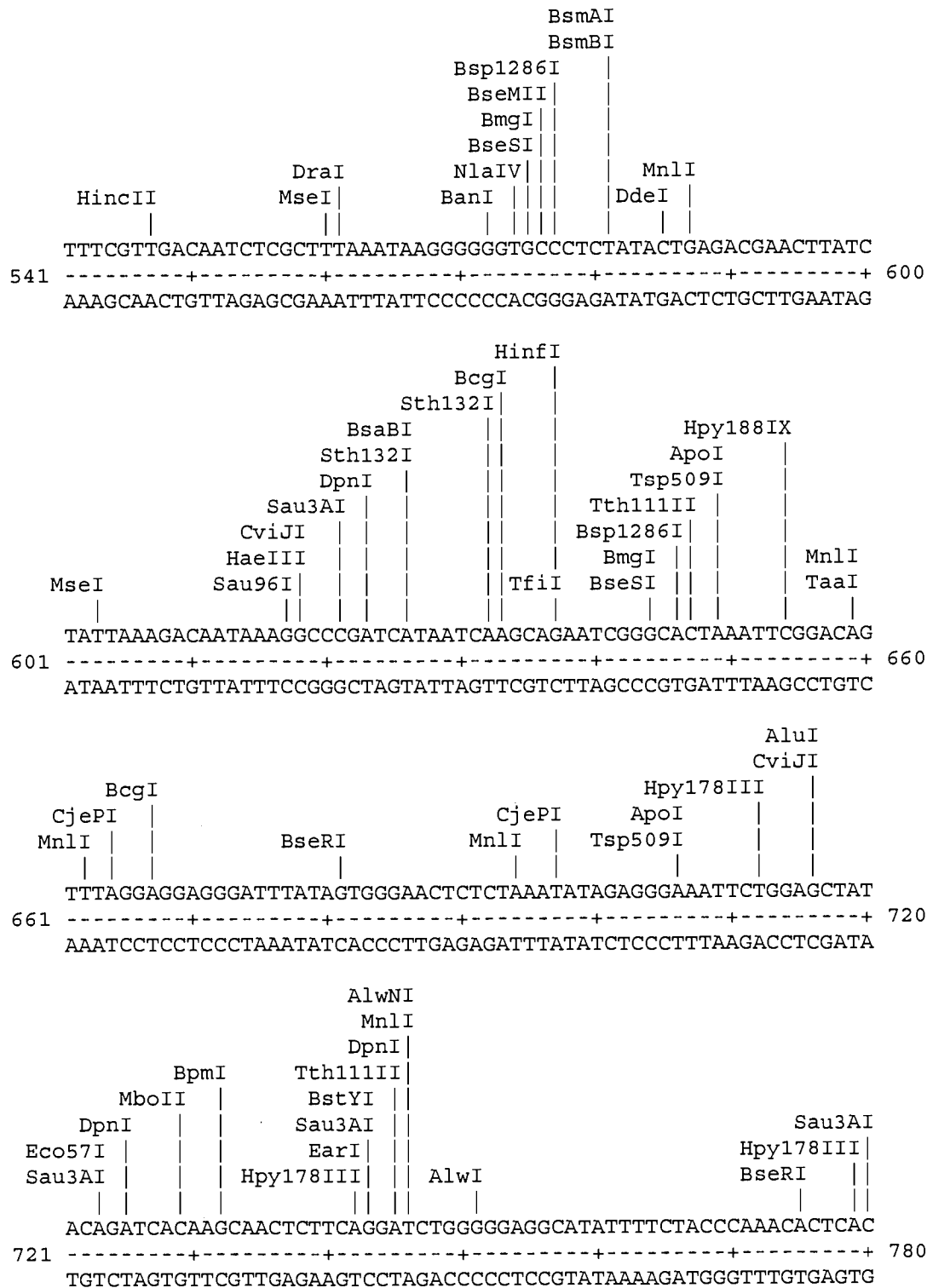


Fig. 24 (con't)

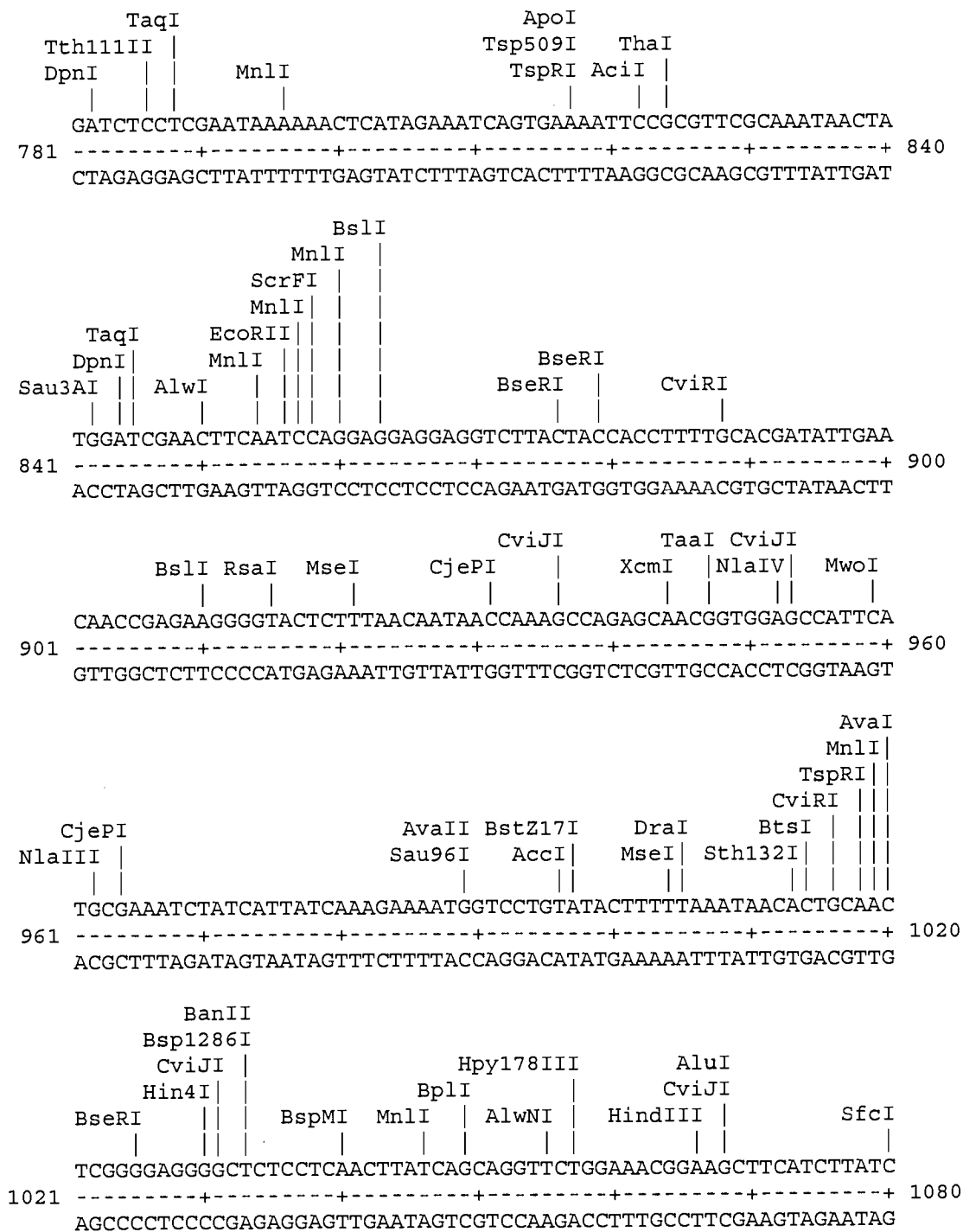


Fig. 24 (con't)

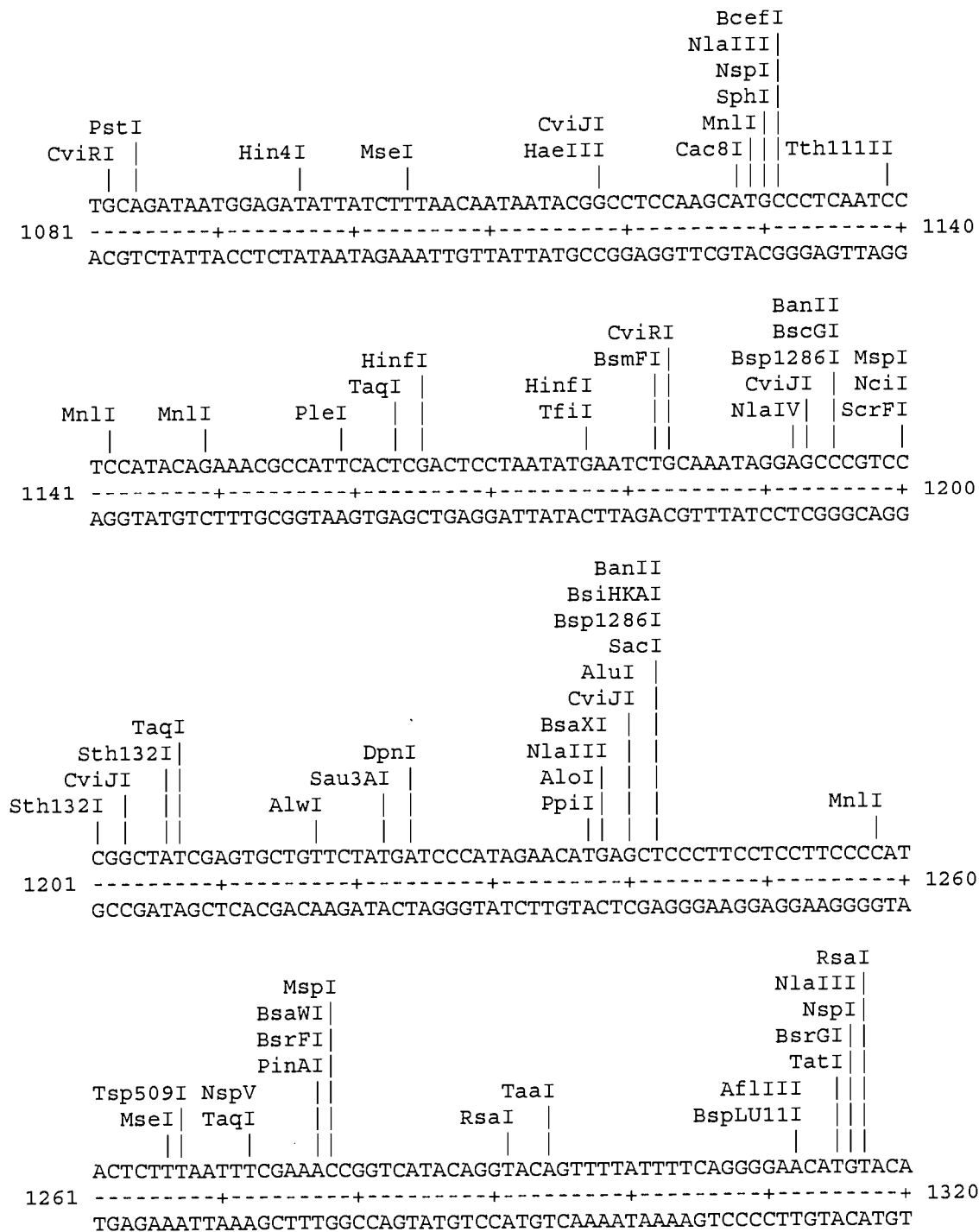


Fig. 24 (con't)

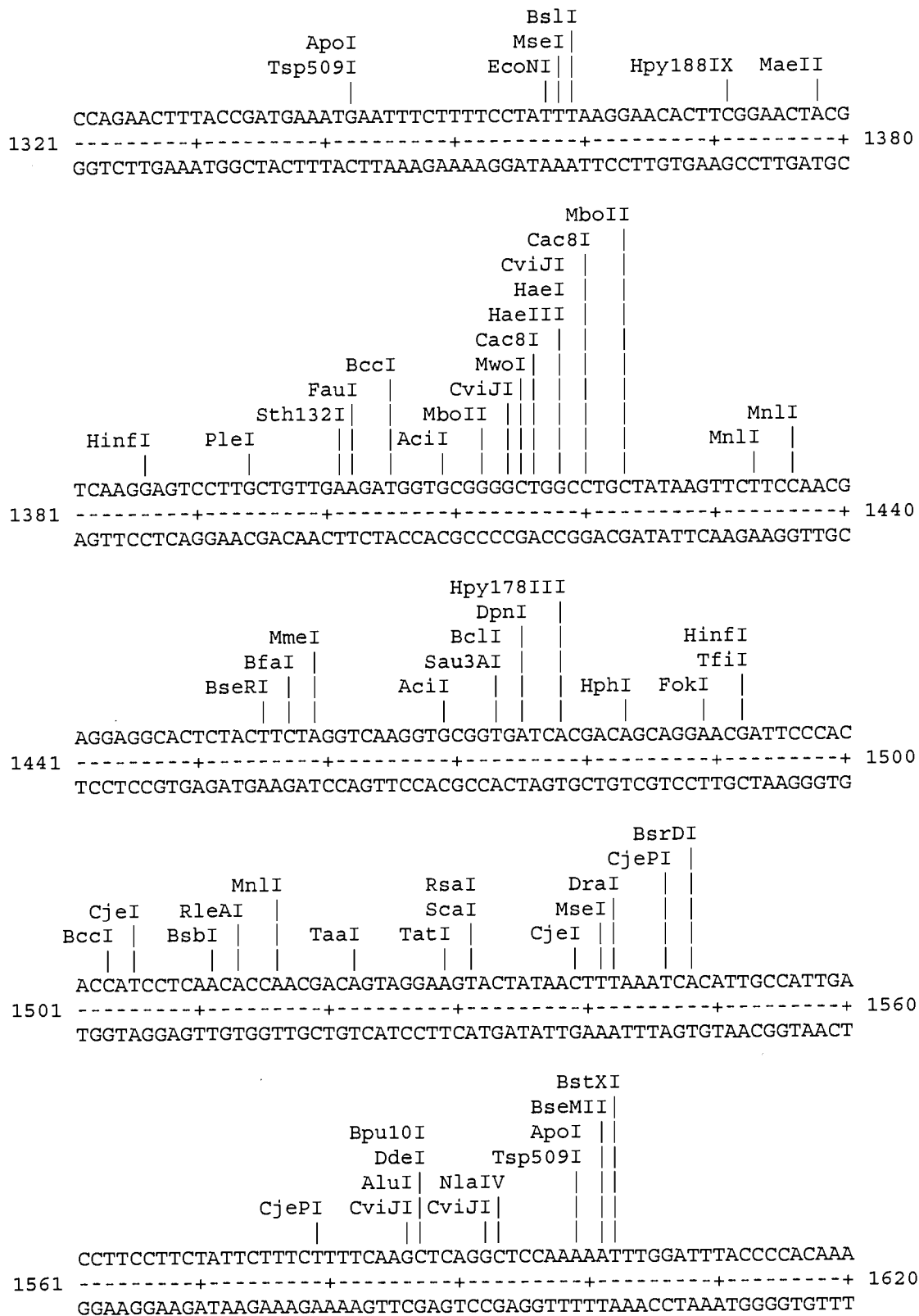


Fig. 24 (con't)

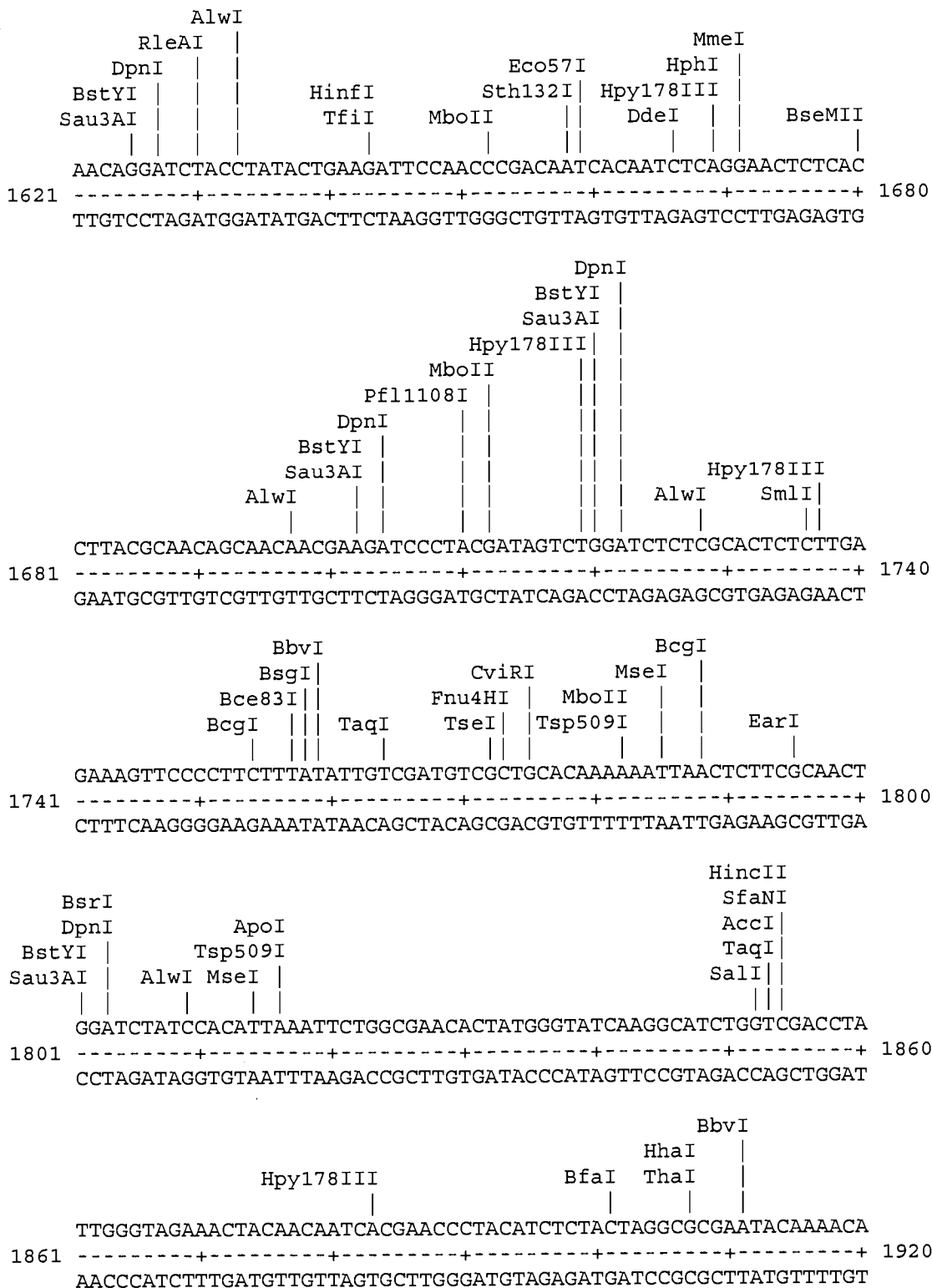


Fig. 24 (con't)

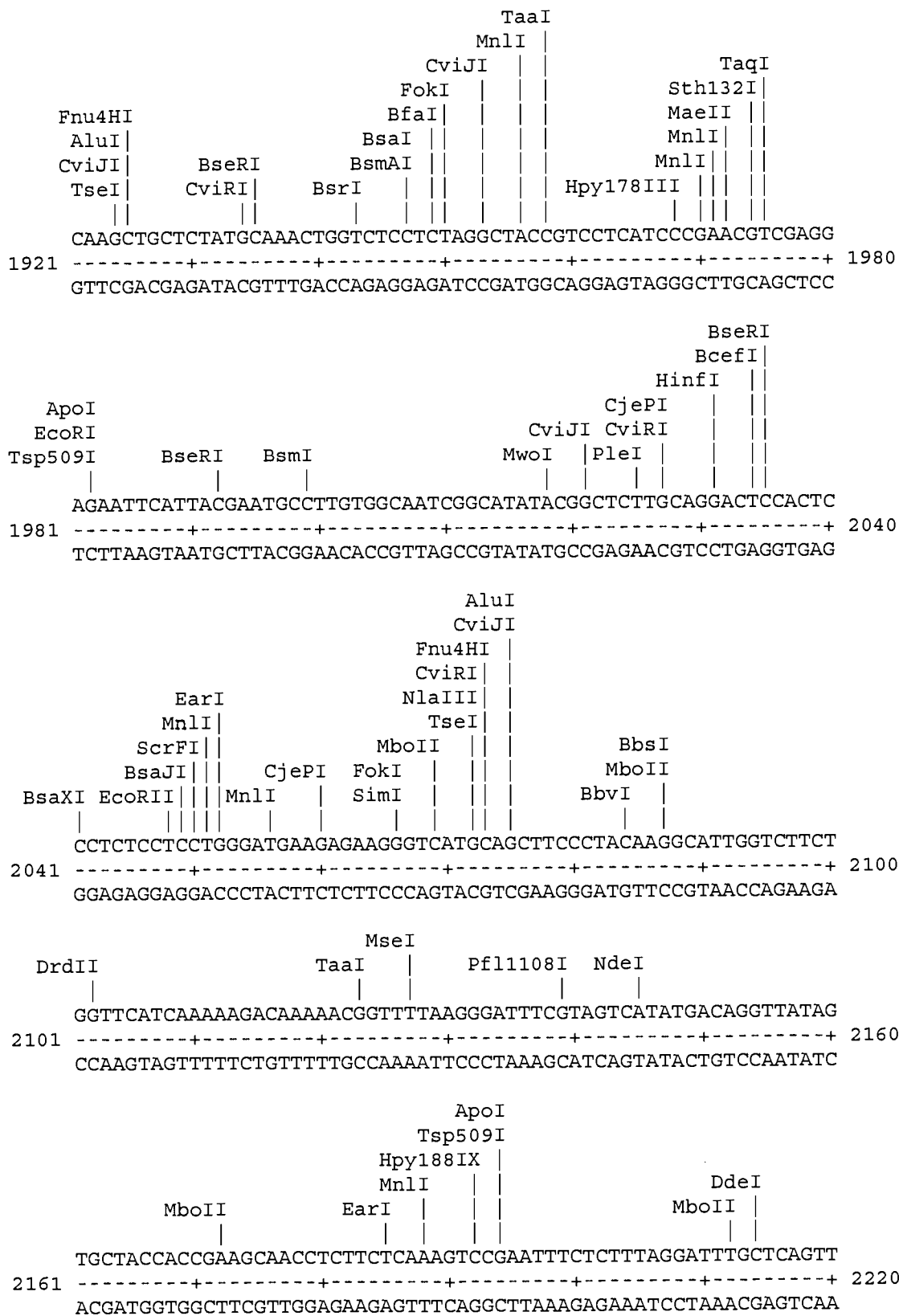




Fig. 24 (con't)

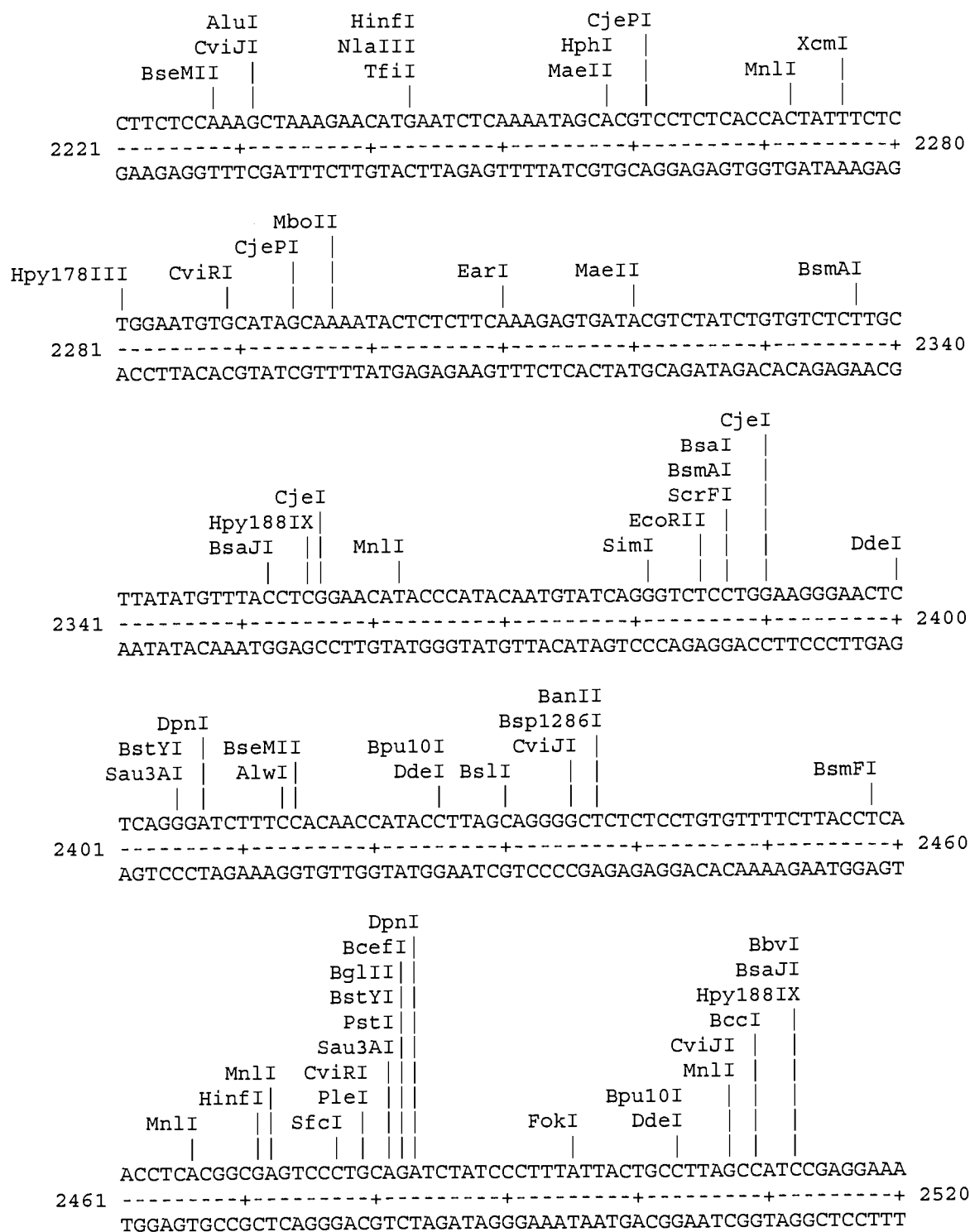
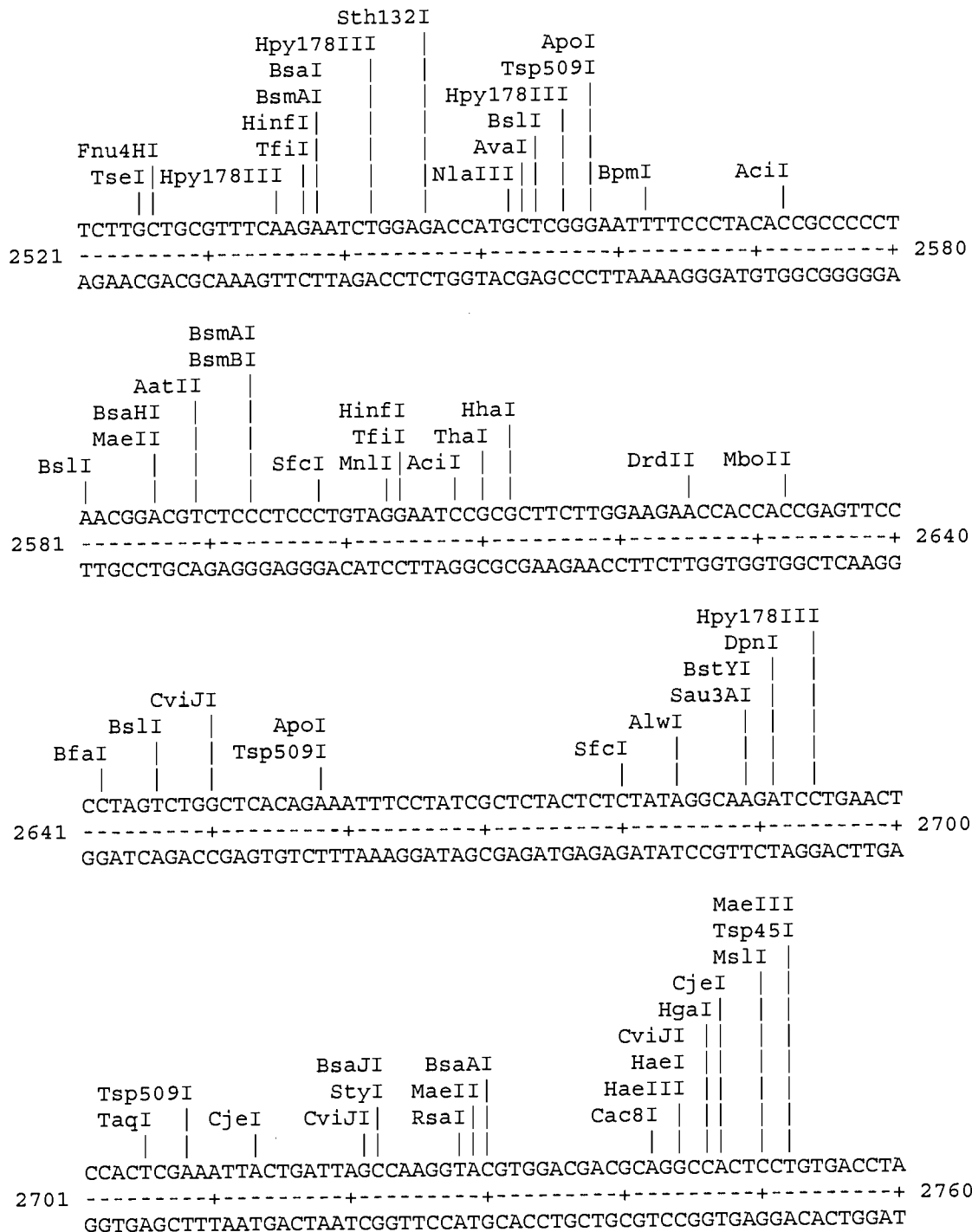


Fig. 24 (con't)



[illegible]

Figure 25:

```

cactgtggat gtgatattcg cagaacctcc cgtcaaatat actctagata taggaagcaa 60

attacgattt taaaccttat ttaacgacag gggtgaggc atg cct ctt tct ttc      114
                                Met Pro Leu Ser Phe
                                1          5

aaa tct tca tct ttt tgt cta ctt gcc tgt tta tgt agt gca agt tgc      162
Lys Ser Ser Ser Phe Cys Leu Leu Ala Cys Leu Cys Ser Ala Ser Cys
                                10          15          20

gcg ttt gct gag act aga ctc gga ggg aac ttt gtt cct cca att acg      210
Ala Phe Ala Glu Thr Arg Leu Gly Gly Asn Phe Val Pro Pro Ile Thr
                                25          30          35

aat cag ggt gaa gag atc tta ctc act tca gat ttt gtt tgt tca aac      258
Asn Gln Gly Glu Glu Ile Leu Leu Thr Ser Asp Phe Val Cys Ser Asn
                                40          45          50

ttc ttg ggg gcg agt ttt tca agt tcc ttt atc aat agt tcc agc aat      306
Phe Leu Gly Ala Ser Phe Ser Ser Ser Phe Ile Asn Ser Ser Ser Asn
                                55          60          65

ctc tcc tta tta ggg aag ggc ctt tcc tta acg ttt acc tct tgt caa      354
Leu Ser Leu Leu Gly Lys Gly Leu Ser Leu Thr Phe Thr Ser Cys Gln
                                70          75          80          85

gct cct aca aat agt aac tat gcg cta ctt tct gcc gca gag act ctg      402
Ala Pro Thr Asn Ser Asn Tyr Ala Leu Leu Ser Ala Ala Glu Thr Leu
                                90          95          100

acc ttc aag aat ttt tct tct ata aac ttt aca ggg aac caa tcg aca      450
Thr Phe Lys Asn Phe Ser Ser Ile Asn Phe Thr Gly Asn Gln Ser Thr
                                105          110          115

gga ctt ggc ggc ctc atc tac gga aaa gat att gtt ttc caa tct atc      498
Gly Leu Gly Gly Leu Ile Tyr Gly Lys Asp Ile Val Phe Gln Ser Ile
                                120          125          130

aaa gat ttg atc ttc act acg aac cgt gtt gcc tat tct cca gca tct      546
Lys Asp Leu Ile Phe Thr Thr Asn Arg Val Ala Tyr Ser Pro Ala Ser
                                135          140          145

gta act acg tcg gca act ccc gca atc act aca gta act aca gga gcc      594
Val Thr Thr Ser Ala Thr Pro Ala Ile Thr Thr Val Thr Thr Gly Ala
                                150          155          160          165

tct gct ctc caa cct aca gac tca ctc act gtc gaa aac ata tcc caa      642
Ser Ala Leu Gln Pro Thr Asp Ser Leu Thr Val Glu Asn Ile Ser Gln
Ser Ala Leu Gln Pro Thr Asp Ser Leu Thr Val Glu Asn Ile Ser Gln
                                170          175          180

tcg atc aag ttc ttt ggg aac ctt gcc aac ttc ggc tct gca att agc      690
Ser Ile Lys Phe Phe Gly Asn Leu Ala Asn Phe Gly Ser Ala Ile Ser
Ser Ile Lys Phe Phe Gly Asn Leu Ala Asn Phe Gly Ser Ala Ile Ser
                                185          190          195

```

Fig. 25 (con't)

agt tct ccc acg gca gtc gtt aaa ttc atc aat aac acc gct acc atg	738
Ser Ser Pro Thr Ala Val Val Lys Phe Ile Asn Asn Thr Ala Thr Met	
Ser Ser Pro Thr Ala Val Val Lys Phe Ile Asn Asn Thr Ala Thr Met	
200 205 210	
agc ttc tcc cat aac ttt act tcg tca gga ggc ggc gtg att tat gga	786
Ser Phe Ser Ser His Asn Phe Thr Ser Ser Gly Gly Gly Val Ile Tyr Gly	
Ser Phe Ser Ser His Asn Phe Thr Ser Ser Gly Gly Gly Val Ile Tyr Gly	
215 220 225	
gga agc tct ctc ctt ttt gaa aac aat tct gga tgc atc atc ttc acc	834
Gly Ser Ser Leu Leu Phe Glu Asn Asn Ser Gly Cys Ile Ile Phe Thr	
Gly Ser Ser Leu Leu Phe Glu Asn Asn Ser Gly Cys Ile Ile Phe Thr	
230 235 240 245	
gcc aac tcc tgt gtg aac agc tta aaa ggc gtc acc cct tca tca gga	882
Ala Asn Ser Cys Val Asn Ser Leu Lys Gly Val Thr Pro Ser Ser Gly	
Ala Asn Ser Cys Val Asn Ser Leu Lys Gly Val Thr Pro Ser Ser Gly	
250 255 260	
acc tat gct tta gga agt ggc gga gcc atc tgc atc cct acg gga act	930
Thr Tyr Ala Leu Gly Ser Gly Gly Ala Ile Cys Ile Pro Thr Gly Thr	
Thr Tyr Ala Leu Gly Ser Gly Gly Ala Ile Cys Ile Pro Thr Gly Thr	
265 270 275	
ttc gaa tta aaa aac aat cag ggg aag tgc acc ttc tct tat aat ggt	978
Phe Glu Leu Lys Asn Asn Gln Gly Lys Cys Thr Phe Ser Tyr Asn Gly	
Phe Glu Leu Lys Asn Asn Gln Gly Lys Cys Thr Phe Ser Tyr Asn Gly	
280 285 290	
aca cca aat gat gcg ggt gcg atc tac gcc gaa acc tgc aac atc gta	1026
Thr Pro Asn Asp Ala Gly Ala Ile Tyr Ala Glu Thr Cys Asn Ile Val	
Thr Pro Asn Asp Ala Gly Ala Ile Tyr Ala Glu Thr Cys Asn Ile Val	
295 300 305	
ggg aac cag ggt gcc ttg ctc cta gat agc aac act gca gcg aga aat	1074
Gly Asn Gln Gly Ala Leu Leu Leu Asp Ser Asn Thr Ala Ala Arg Asn	
Gly Asn Gln Gly Ala Leu Leu Leu Asp Ser Asn Thr Ala Ala Arg Asn	
310 315 320 325	
ggc gga gcc atc tgt gct aaa gtg ctc aat att caa gga cgc ggt cct	1122
Gly Gly Ala Ile Cys Ala Lys Val Leu Asn Ile Gln Gly Arg Gly Pro	
Gly Gly Ala Ile Cys Ala Lys Val Leu Asn Ile Gln Gly Arg Gly Pro	
330 335 340	
att gaa ttc tct aga aac cgc gcg gag aag ggt gga gct att ttc ata	1170
Ile Glu Phe Ser Arg Asn Arg Ala Glu Lys Gly Gly Ala Ile Phe Ile	
Ile Glu Phe Ser Arg Asn Arg Ala Glu Lys Gly Gly Ala Ile Phe Ile	
345 350 355	
ggc ccc tct gtt gga gac cct gcg aag caa aca tcg aca ctt acg att	1218
Gly Pro Ser Val Gly Asp Pro Ala Lys Gln Thr Ser Thr Leu Thr Ile	
Gly Pro Ser Val Gly Asp Pro Ala Lys Gln Thr Ser Thr Leu Thr Ile	
360 365 370	
ttg gct tcc gaa ggt gat att gcg ttc caa gga aac atg ctc aat aca	1266
Leu Ala Ser Glu Gly Asp Ile Ala Phe Gln Gly Asn Met Leu Asn Thr	
Leu Ala Ser Glu Gly Asp Ile Ala Phe Gln Gly Asn Met Leu Asn Thr	
375 380 385	

Fig. 25 (con't)

aaa cct gga atc cgc aat gcc atc act gta gaa gca ggg gga gag att	1314
Lys Pro Gly Ile Arg Asn Ala Ile Thr Val Glu Ala Gly Gly Glu Ile	
Lys Pro Gly Ile Arg Asn Ala Ile Thr Val Glu Ala Gly Gly Glu Ile	
390 395 400 405	
gtg tct cta tct gca caa gga ggc tca cgt ctt gta ttt tat gat ccc	1362
Val Ser Leu Ser Ala Gln Gly Gly Ser Arg Leu Val Phe Tyr Asp Pro	
Val Ser Leu Ser Ala Gln Gly Gly Ser Arg Leu Val Phe Tyr Asp Pro	
410 415 420	
att aca cat agc ctc cca acc aca agt ccg tct aat aaa gac att aca	1410
Ile Thr His Ser Leu Pro Thr Thr Ser Pro Ser Asn Lys Asp Ile Thr	
Ile Thr His Ser Leu Pro Thr Thr Ser Pro Ser Asn Lys Asp Ile Thr	
425 430 435	
atc aac gct aat ggc gct tca gga tct gta gtc ttt aca agt aag gga	1458
Ile Asn Ala Asn Gly Ala Ser Gly Ser Val Val Phe Thr Ser Lys Gly	
Ile Asn Ala Asn Gly Ala Ser Gly Ser Val Val Phe Thr Ser Lys Gly	
440 445 450	
ctc tcc tct aca gaa ctc ctg ttg cct gcc aac acg aca act ata ctt	1506
Leu Ser Ser Thr Glu Leu Leu Leu Pro Ala Asn Thr Thr Thr Ile Leu	
Leu Ser Ser Thr Glu Leu Leu Leu Pro Ala Asn Thr Thr Thr Ile Leu	
455 460 465	
cta gga aca gtc aag atc gct agt gga gaa ctg aag att act gac aat	1554
Leu Gly Thr Val Lys Ile Ala Ser Gly Glu Leu Lys Ile Thr Asp Asn	
Leu Gly Thr Val Lys Ile Ala Ser Gly Glu Leu Lys Ile Thr Asp Asn	
470 475 480 485	
gcg gtt gtc aat gtt gct ggc ttc gct act cag ggc tca ggt cag ctt	1602
Ala Val Val Asn Val Ala Gly Phe Ala Thr Gln Gly Ser Gly Gln Leu	
Ala Val Val Asn Val Ala Gly Phe Ala Thr Gln Gly Ser Gly Gln Leu	
490 495 500	
acc ctg ggc tct gga gga acc tta ggg ctg gca aca ccc acg gga gca	1650
Thr Leu Gly Ser Gly Gly Thr Leu Gly Leu Ala Thr Pro Thr Gly Ala	
Thr Leu Gly Ser Gly Gly Thr Leu Gly Leu Ala Thr Pro Thr Gly Ala	
505 510 515	
cct gcc gct gta gac ttt acg att gga aag tta gca ttc gat cct ttt	1698
Pro Ala Ala Val Asp Phe Thr Ile Gly Lys Leu Ala Phe Asp Pro Phe	
Pro Ala Ala Val Asp Phe Thr Ile Gly Lys Leu Ala Phe Asp Pro Phe	
520 525 530	
tcc ttc cta aaa aga gat ttt gtt tca gca tca gta aat gca ggc aca	1746
Ser Phe Leu Lys Arg Asp Phe Val Ser Ala Ser Val Asn Ala Gly Thr	
Ser Phe Leu Lys Arg Asp Phe Val Ser Ala Ser Val Asn Ala Gly Thr	
535 540 545	
aaa aac gtc act tta aca gga gct ctg gtt ctt gat gaa cat gac gtt	1794
Lys Asn Val Thr Leu Thr Gly Ala Leu Val Leu Asp Glu His Asp Val	
Lys Asn Val Thr Leu Thr Gly Ala Leu Val Leu Asp Glu His Asp Val	
550 555 560 565	
aca gat ctt tat gat atg gtg tca tta caa tct cca gta gca att cct	1842
Thr Asp Leu Tyr Asp Met Val Ser Leu Gln Ser Pro Val Ala Ile Pro	
Thr Asp Leu Tyr Asp Met Val Ser Leu Gln Ser Pro Val Ala Ile Pro	
570 575 580	

Fig. 25 (con't)

atc gct gtt ttc aaa gga gca acc gtt act aag aca gga ttt cct gat	1890
Ile Ala Val Phe Lys Gly Ala Thr Val Thr Lys Thr Gly Phe Pro Asp	
Ile Ala Val Phe Lys Gly Ala Thr Val Thr Lys Thr Gly Phe Pro Asp	
585 590 595	
ggg gag att gcg act cca agc cac tac ggc tac caa gga aag tgg tcc	1938
Gly Glu Ile Ala Thr Pro Ser His Tyr Gly Tyr Gln Gly Lys Trp Ser	
Gly Glu Ile Ala Thr Pro Ser His Tyr Gly Tyr Gln Gly Lys Trp Ser	
600 605 610	
tac aca tgg tcc cgt ccc ctg tta att cca gct cct gat gga gga ttt	1986
Tyr Thr Trp Ser Arg Pro Leu Leu Ile Pro Ala Pro Asp Gly Gly Phe	
Tyr Thr Trp Ser Arg Pro Leu Leu Ile Pro Ala Pro Asp Gly Gly Phe	
615 620 625	
cct gga ggt ccc tct cct agc gca aat act ctc tat gct gta tgg aat	2034
Pro Gly Gly Pro Ser Pro Ser Ala Asn Thr Leu Tyr Ala Val Trp Asn	
Pro Gly Gly Pro Ser Pro Ser Ala Asn Thr Leu Tyr Ala Val Trp Asn	
630 635 640 645	
tca gac act ctc gtg cgt tct acc tat atc tta gat ccc gag cgt tac	2082
Ser Asp Thr Leu Val Arg Ser Thr Tyr Ile Leu Asp Pro Glu Arg Tyr	
Ser Asp Thr Leu Val Arg Ser Thr Tyr Ile Leu Asp Pro Glu Arg Tyr	
650 655 660	
gga gaa att gtc agc aac agc tta tgg att tcc ttc tta gga aat cag	2130
Gly Glu Ile Val Ser Asn Ser Leu Trp Ile Ser Phe Leu Gly Asn Gln	
Gly Glu Ile Val Ser Asn Ser Leu Trp Ile Ser Phe Leu Gly Asn Gln	
665 670 675	
gca ttc tct gat att ctc caa gat gtt ctt ttg ata gat cat ccc ggg	2178
Ala Phe Ser Asp Ile Leu Gln Asp Val Leu Leu Ile Asp His Pro Gly	
Ala Phe Ser Asp Ile Leu Gln Asp Val Leu Leu Ile Asp His Pro Gly	
680 685 690	
ttg tcc ata acc gcg aaa gct tta gga gcc tat gtc gaa cac aca cca	2226
Leu Ser Ile Thr Ala Lys Ala Leu Gly Ala Tyr Val Glu His Thr Pro	
Leu Ser Ile Thr Ala Lys Ala Leu Gly Ala Tyr Val Glu His Thr Pro	
695 700 705	
aga caa gga cat gag ggc ttt tca ggt cgc tat gga ggc tac caa gct	2274
Arg Gln Gly His Glu Gly Phe Ser Gly Arg Tyr Gly Gly Tyr Gln Ala	
Arg Gln Gly His Glu Gly Phe Ser Gly Arg Tyr Gly Gly Tyr Gln Ala	
710 715 720 725	
gcg cta tct atg aac tac acg gac cac act acg tta gga ctt tct ttc	2322
Ala Leu Ser Met Asn Tyr Thr Asp His Thr Thr Leu Gly Leu Ser Phe	
Ala Leu Ser Met Asn Tyr Thr Asp His Thr Thr Leu Gly Leu Ser Phe	
730 735 740	
ggg cag ctt tat gga aaa act aac gcc aac ccc tac gat tca cgt tgc	2370
Gly Gln Leu Tyr Gly Lys Thr Asn Ala Asn Pro Tyr Asp Ser Arg Cys	
Gly Gln Leu Tyr Gly Lys Thr Asn Ala Asn Pro Tyr Asp Ser Arg Cys	
745 750 755	
tca gaa caa atg tat tta ctc tcg ttc ttt ggt caa ttc cct atc gtg	2418
Ser Glu Gln Met Tyr Leu Leu Ser Phe Phe Gly Gln Phe Pro Ile Val	
Ser Glu Gln Met Tyr Leu Leu Ser Phe Phe Phe Gly Gln Phe Pro Ile Val	
760 765 770	

Fig. 25 (con't)

act	caa	aag	agc	gag	gcc	tta	att	tcc	tgg	aaa	gca	gct	tat	ggg	tat	2466
Thr	Gln	Lys	Ser	Glu	Ala	Leu	Ile	Ser	Trp	Lys	Ala	Ala	Tyr	Gly	Tyr	
Thr	Gln	Lys	Ser	Glu	Ala	Leu	Ile	Ser	Trp	Lys	Ala	Ala	Tyr	Gly	Tyr	
	775					780					785					
tcc	aaa	aat	cac	cta	aat	acc	acc	tac	ctc	aga	cct	gac	aaa	gct	cca	2514
Ser	Lys	Asn	His	Leu	Asn	Thr	Thr	Tyr	Leu	Arg	Pro	Asp	Lys	Ala	Pro	
Ser	Lys	Asn	His	Leu	Asn	Thr	Thr	Tyr	Leu	Arg	Pro	Asp	Lys	Ala	Pro	
	790				795					800					805	
aaa	tct	caa	ggg	caa	tgg	cat	aac	aat	agt	tac	tat	ggt	ctt	att	tct	2562
Lys	Ser	Gln	Gly	Gln	Trp	His	Asn	Asn	Ser	Tyr	Tyr	Val	Leu	Ile	Ser	
Lys	Ser	Gln	Gly	Gln	Trp	His	Asn	Asn	Ser	Tyr	Tyr	Val	Leu	Ile	Ser	
				810					815					820		
gca	gaa	cat	cct	ttc	cta	aac	tgg	tgt	ctt	ctt	aca	aga	cct	ctg	gct	2610
Ala	Glu	His	Pro	Phe	Leu	Asn	Trp	Cys	Leu	Leu	Thr	Arg	Pro	Leu	Ala	
Ala	Glu	His	Pro	Phe	Leu	Asn	Trp	Cys	Leu	Leu	Thr	Arg	Pro	Leu	Ala	
			825					830					835			
caa	gct	tgg	gat	ctt	tca	ggg	ttt	att	tcc	gca	gaa	ttc	cta	ggg	ggg	2658
Gln	Ala	Trp	Asp	Leu	Ser	Gly	Phe	Ile	Ser	Ala	Glu	Phe	Leu	Gly	Gly	
Gln	Ala	Trp	Asp	Leu	Ser	Gly	Phe	Ile	Ser	Ala	Glu	Phe	Leu	Gly	Gly	
	840						845					850				
tgg	caa	agt	aag	ttc	aca	gaa	act	gga	gat	ctg	caa	cgt	agc	ttt	agt	2706
Trp	Gln	Ser	Lys	Phe	Thr	Glu	Thr	Gly	Asp	Leu	Gln	Arg	Ser	Phe	Ser	
Trp	Gln	Ser	Lys	Phe	Thr	Glu	Thr	Gly	Asp	Leu	Gln	Arg	Ser	Phe	Ser	
	855					860					865					
aga	ggg	aaa	ggg	tac	aat	gtt	tcc	cta	ccg	ata	gga	tgt	tct	tct	caa	2754
Arg	Gly	Lys	Gly	Tyr	Asn	Val	Ser	Leu	Pro	Ile	Gly	Cys	Ser	Ser	Gln	
Arg	Gly	Lys	Gly	Tyr	Asn	Val	Ser	Leu	Pro	Ile	Gly	Cys	Ser	Ser	Gln	
	870				875				880						885	
tgg	ttc	aca	cca	ttt	aag	aag	gct	cct	tct	aca	ctg	acc	atc	aaa	ctt	2802
Trp	Phe	Thr	Pro	Phe	Lys	Lys	Ala	Pro	Ser	Thr	Leu	Thr	Ile	Lys	Leu	
Trp	Phe	Thr	Pro	Phe	Lys	Lys	Ala	Pro	Ser	Thr	Leu	Thr	Ile	Lys	Leu	
				890					895					900		
gcc	tac	aag	cct	gat	atc	tat	cgt	gtc	aac	cct	cac	aat	att	gtg	act	2850
Ala	Tyr	Lys	Pro	Asp	Ile	Tyr	Arg	Val	Asn	Pro	His	Asn	Ile	Val	Thr	
Ala	Tyr	Lys	Pro	Asp	Ile	Tyr	Arg	Val	Asn	Pro	His	Asn	Ile	Val	Thr	
			905					910					915			
gtc	gtc	tca	aac	caa	gag	agc	act	tcg	atc	tca	gga	gca	aat	cta	cgc	2898
Val	Val	Ser	Asn	Gln	Glu	Ser	Thr	Ser	Ile	Ser	Gly	Ala	Asn	Leu	Arg	
Val	Val	Ser	Asn	Gln	Glu	Ser	Thr	Ser	Ile	Ser	Gly	Ala	Asn	Leu	Arg	
			920				925					930				
cgc	cac	ggg	ttg	ttt	gta	caa	atc	cat	gat	gta	gta	gat	ctc	acc	gag	2946
Arg	His	Gly	Leu	Phe	Val	Gln	Ile	His	Asp	Val	Val	Asp	Leu	Thr	Glu	
Arg	His	Gly	Leu	Phe	Val	Gln	Ile	His	Asp	Val	Val	Asp	Leu	Thr	Glu	
	935					940					945					



Fig. 25 (con't)

```

gac act cag gcc ttt cta aac tat acc ttt gac ggg aaa aat gga ttt 2994
Asp Thr Gln Ala Phe Leu Asn Tyr Thr Phe Asp Gly Lys Asn Gly Phe
Asp Thr Gln Ala Phe Leu Asn Tyr Thr Phe Asp Gly Lys Asn Gly Phe
950                      955                      960                      965

aca aac cac cga gtg tct aca gga cta aaa tcc aca ttt taaaactcta 3043
Thr Asn His Arg Val Ser Thr Gly Leu Lys Ser Thr Phe
Thr Asn His Arg Val Ser Thr Gly Leu Lys Ser Thr Phe
                      970                      975

agctctgctt agagttttct gtagccccgg tcgtcttaga atcctctatc catcatcgaa 3103

gaacttagca atgaaggcca agattctcac tctatgagaa cccccc 3150

```

Figure 26 (RY-47)

## Restriction enzyme analysis of CPN100630

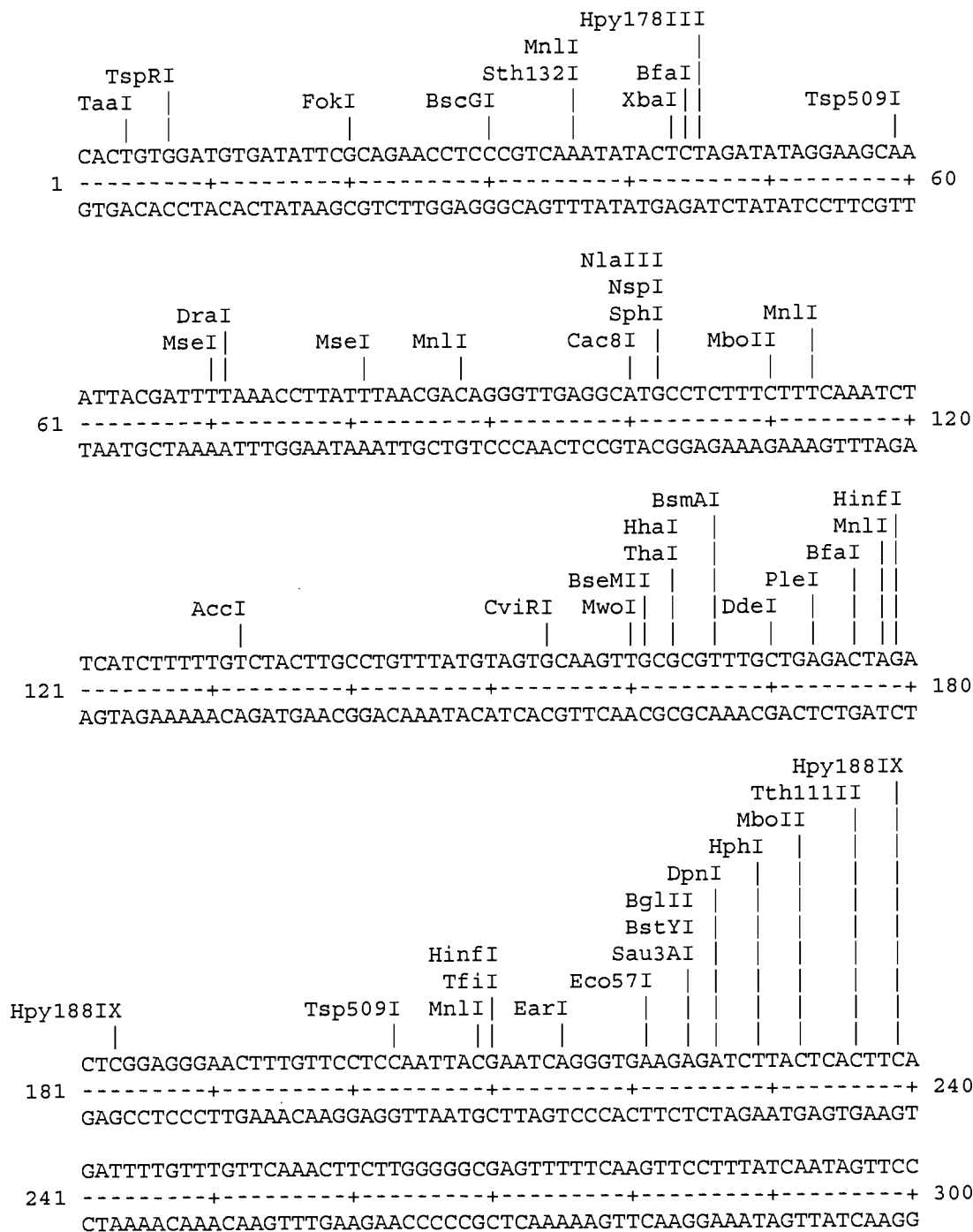


Fig. 26 (con't)

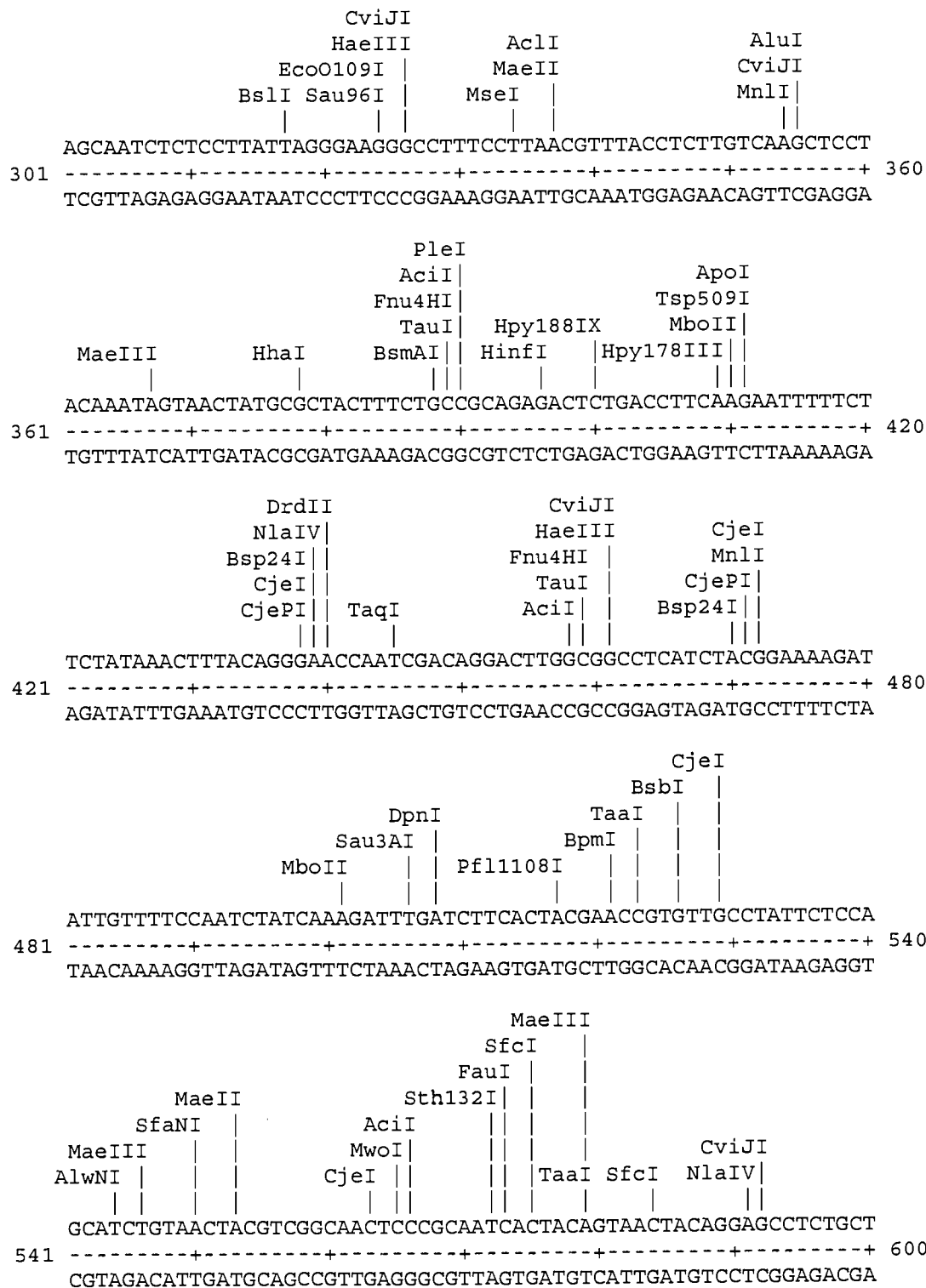


Fig. 26 (con't)

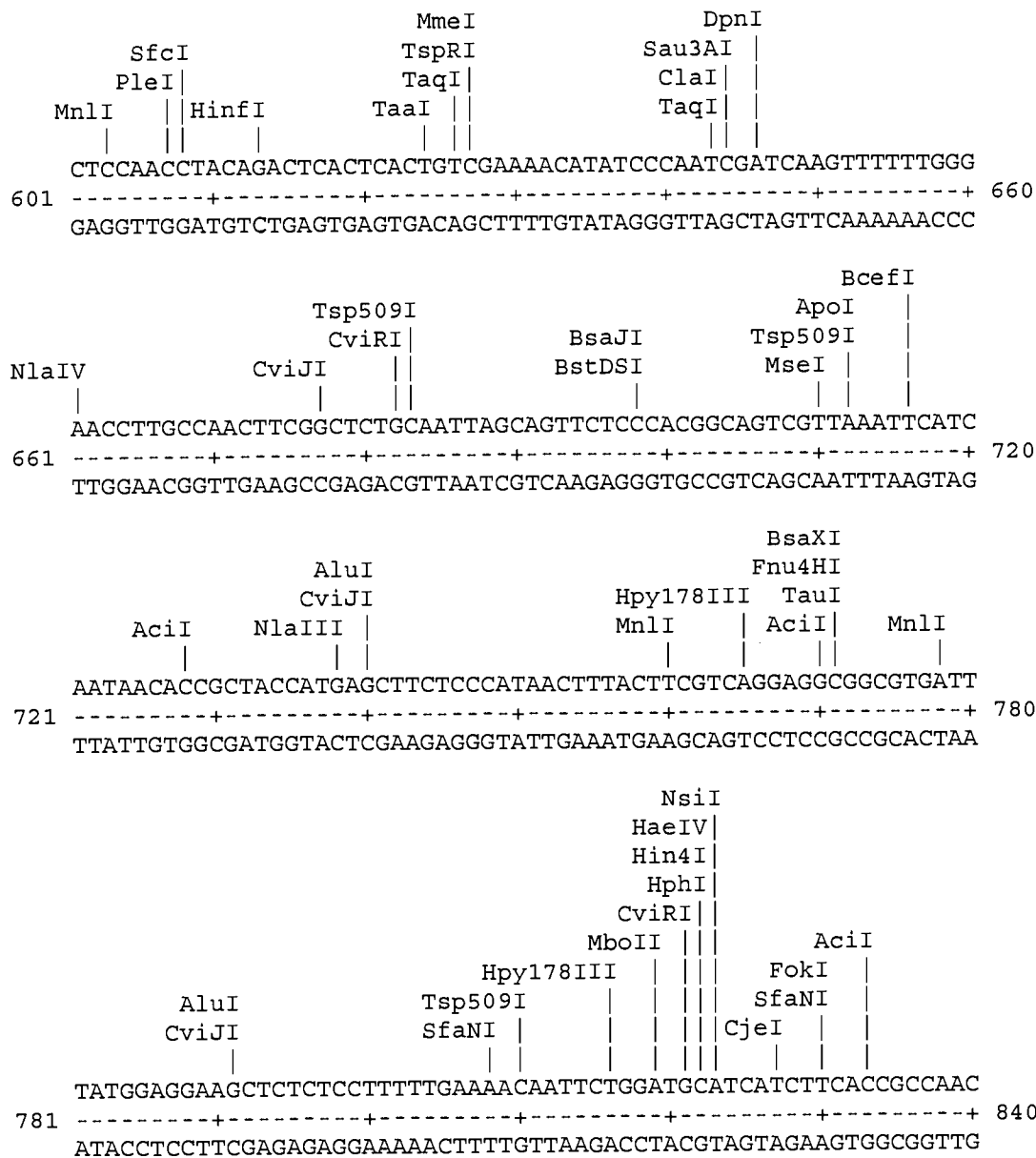
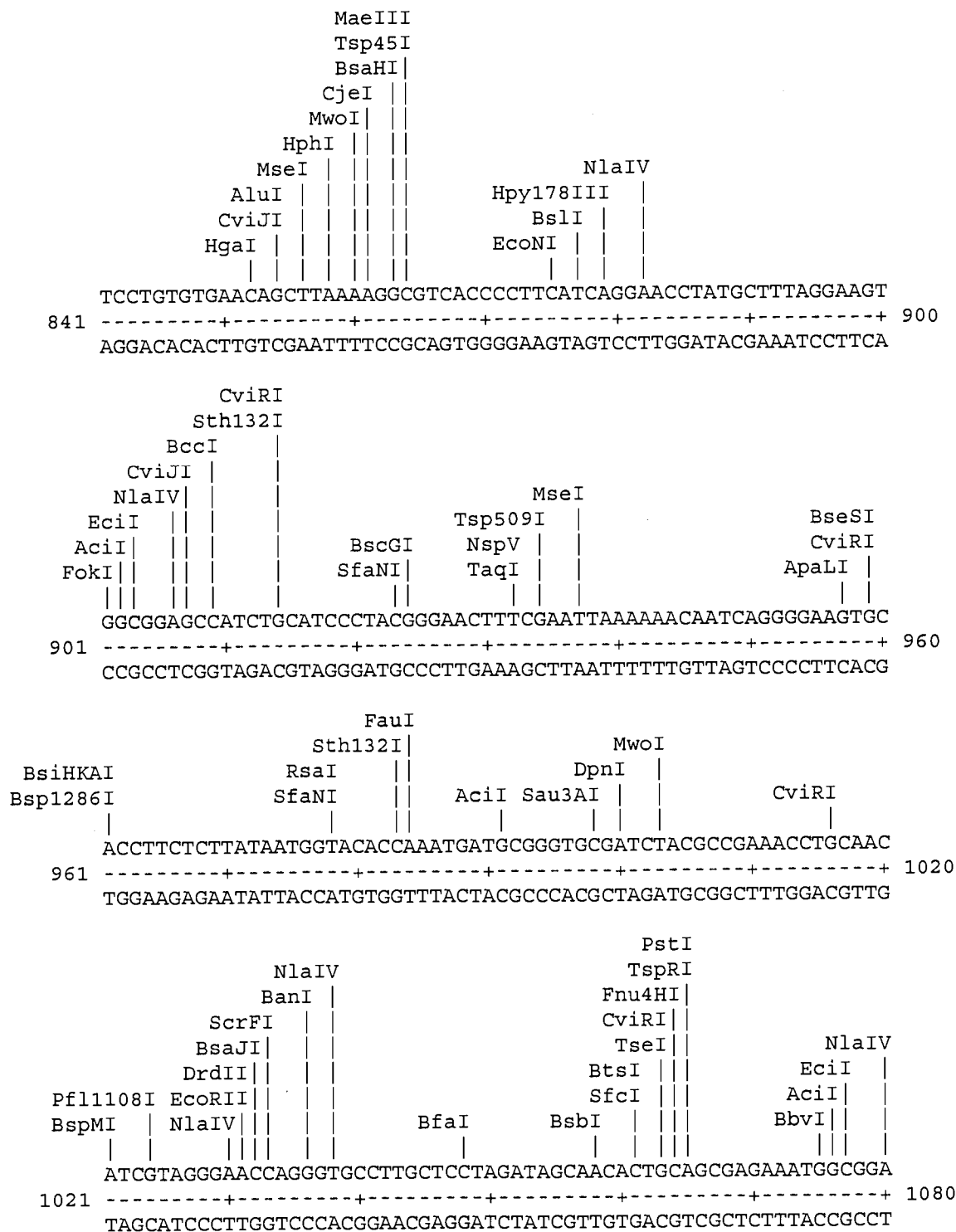


Fig. 26 (con't)



[illegible]

Fig. 26 (con't)

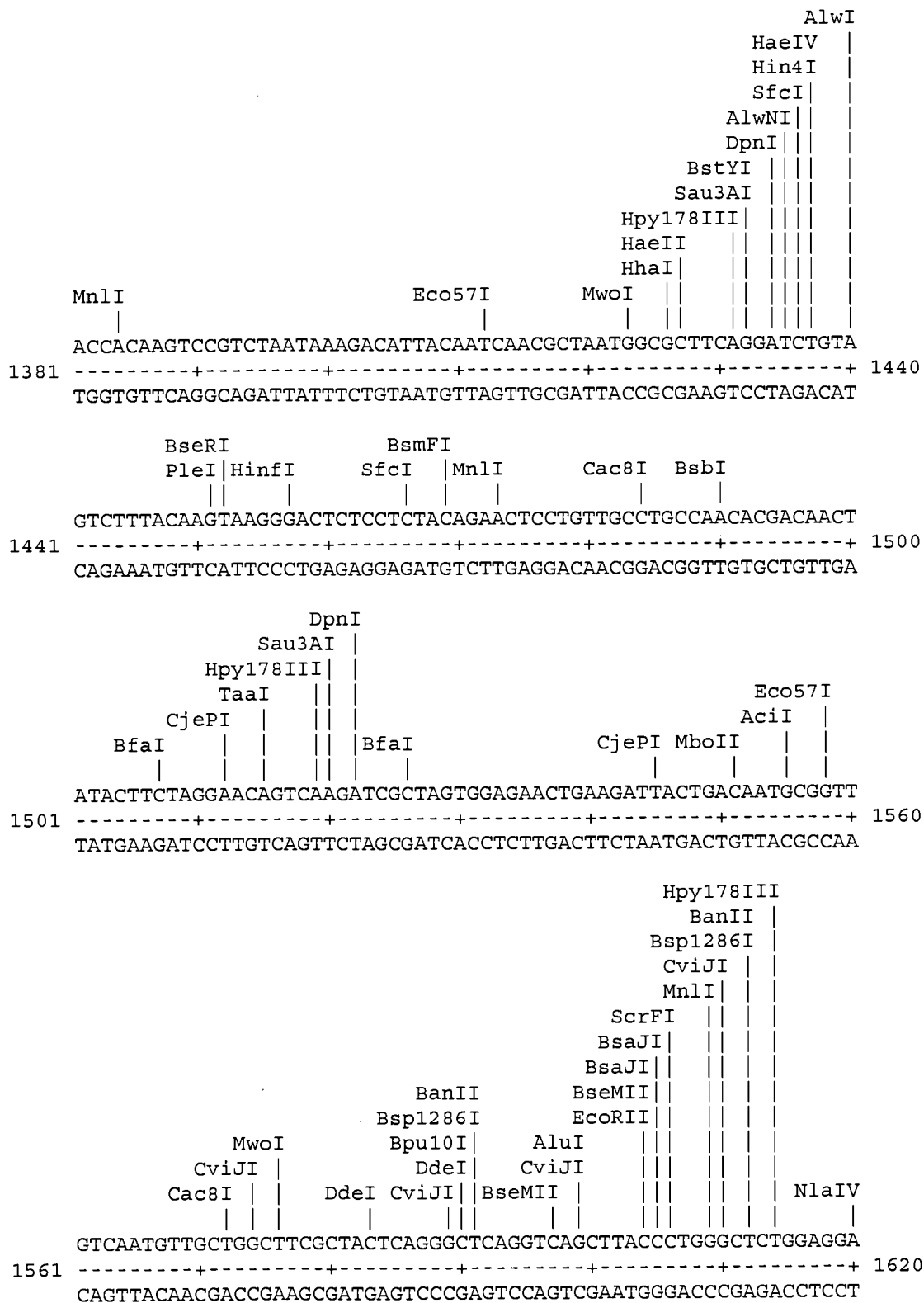


Fig. 26 (con't)

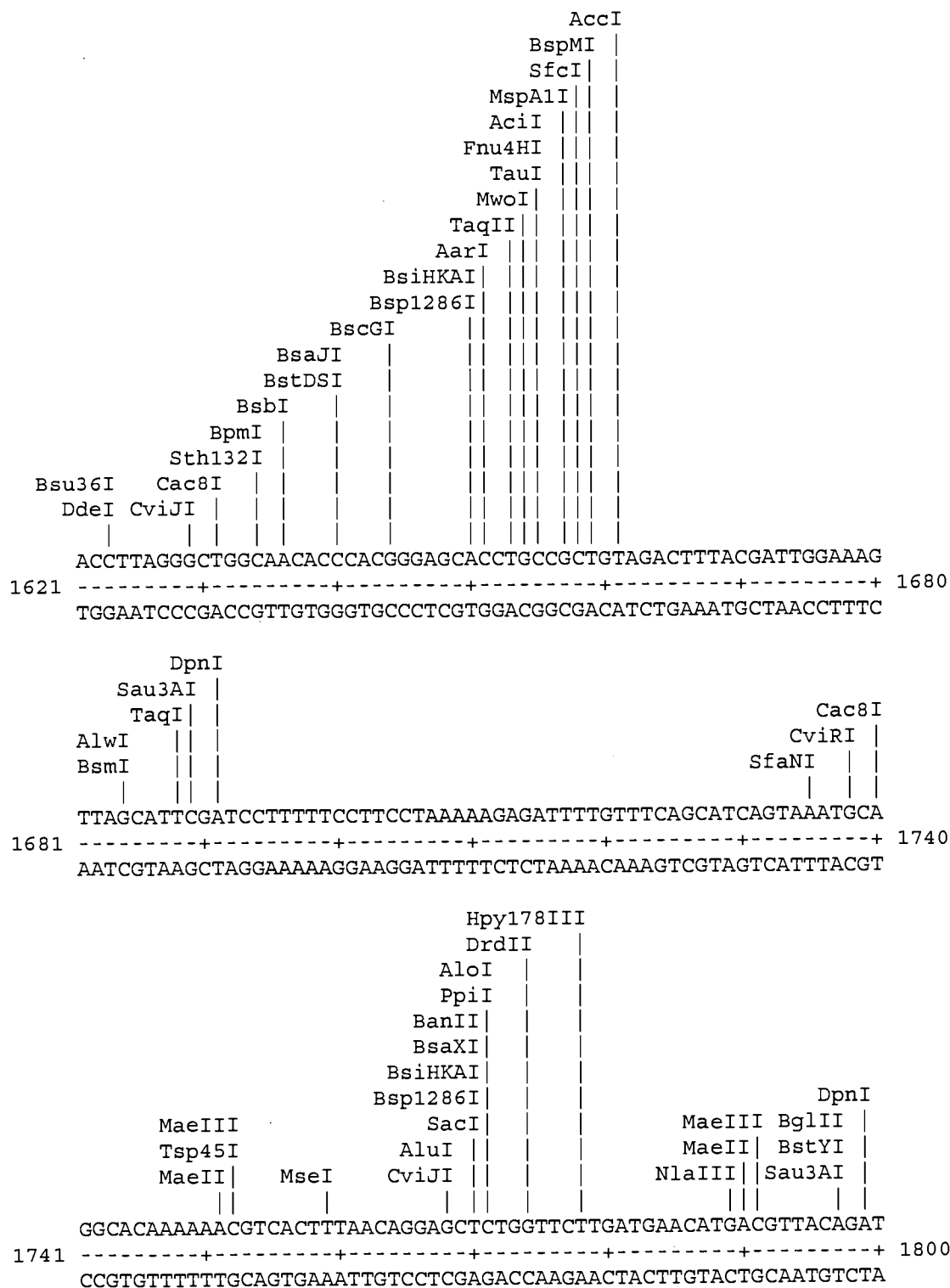




Fig. 26 (con't)

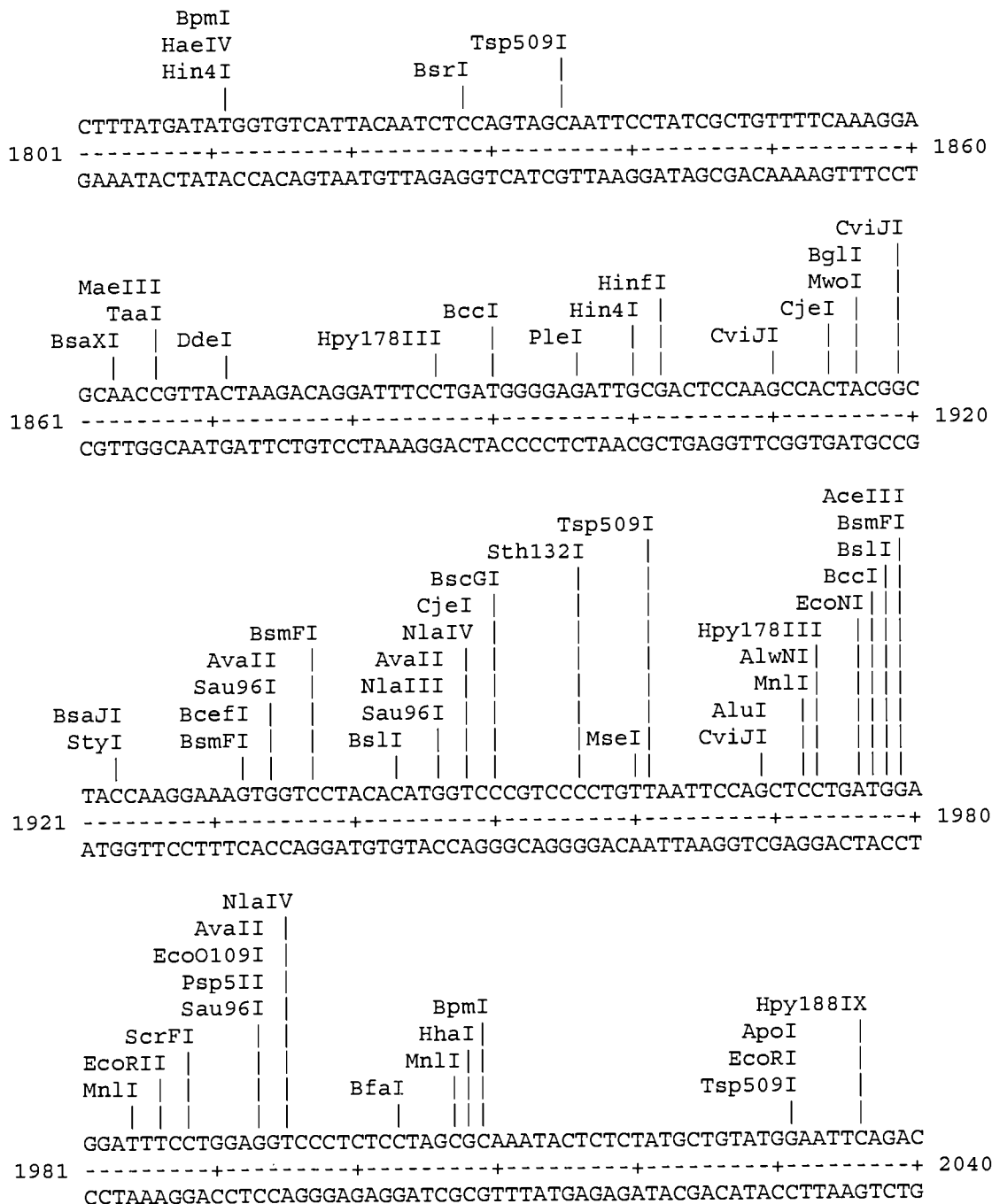


Fig. 26 (con't)

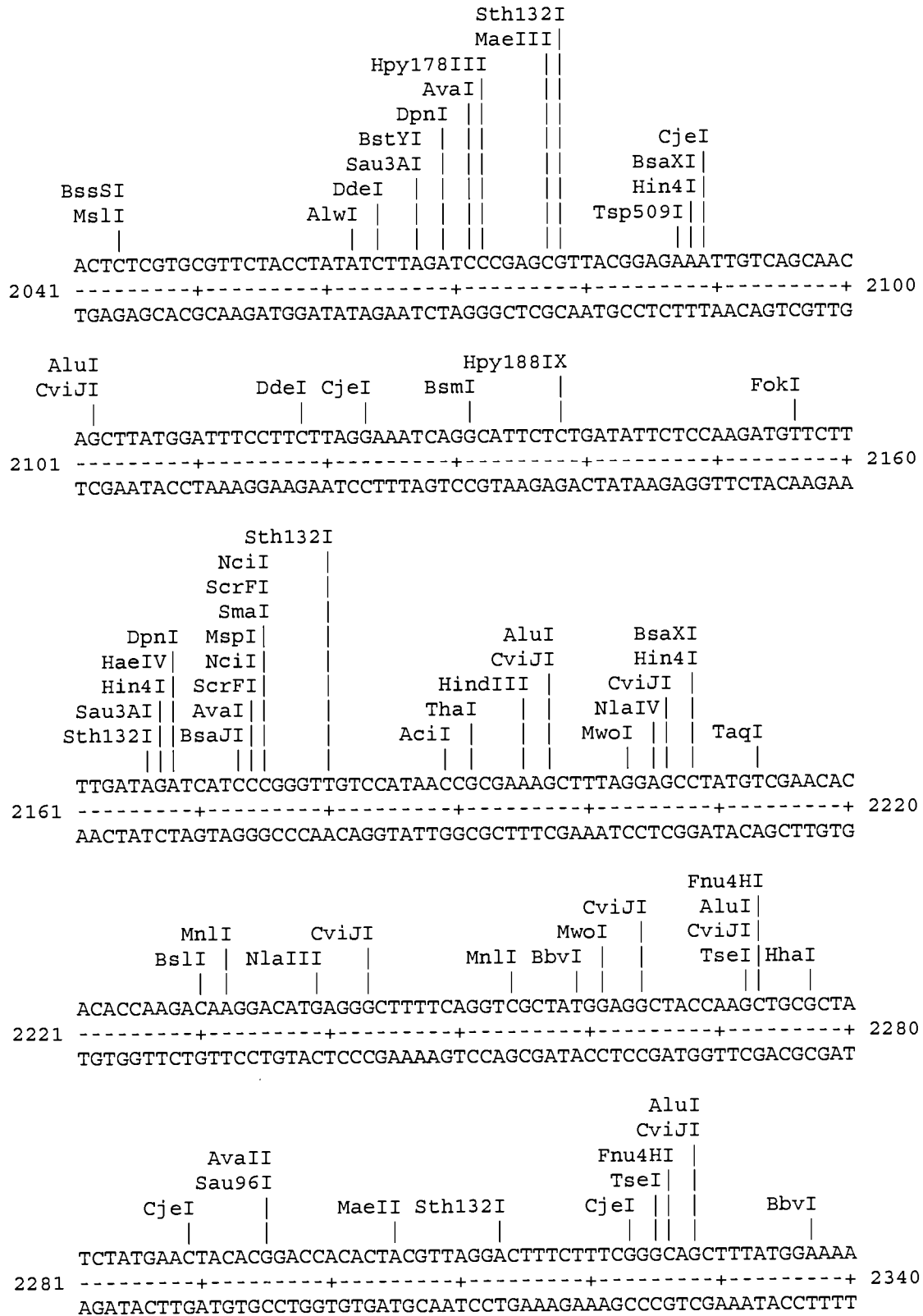


Fig. 26 (con't)

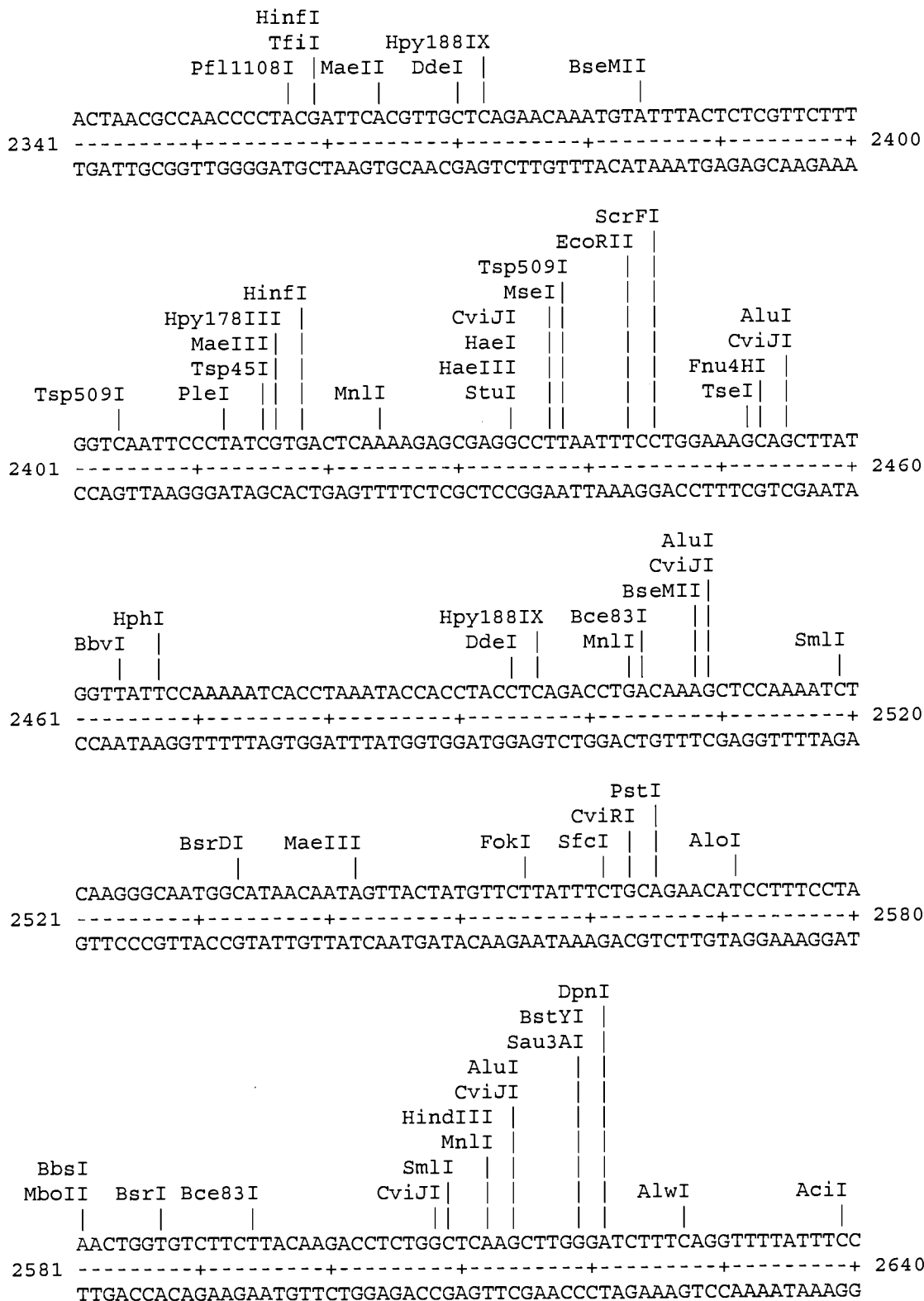


Fig. 26 (con't)

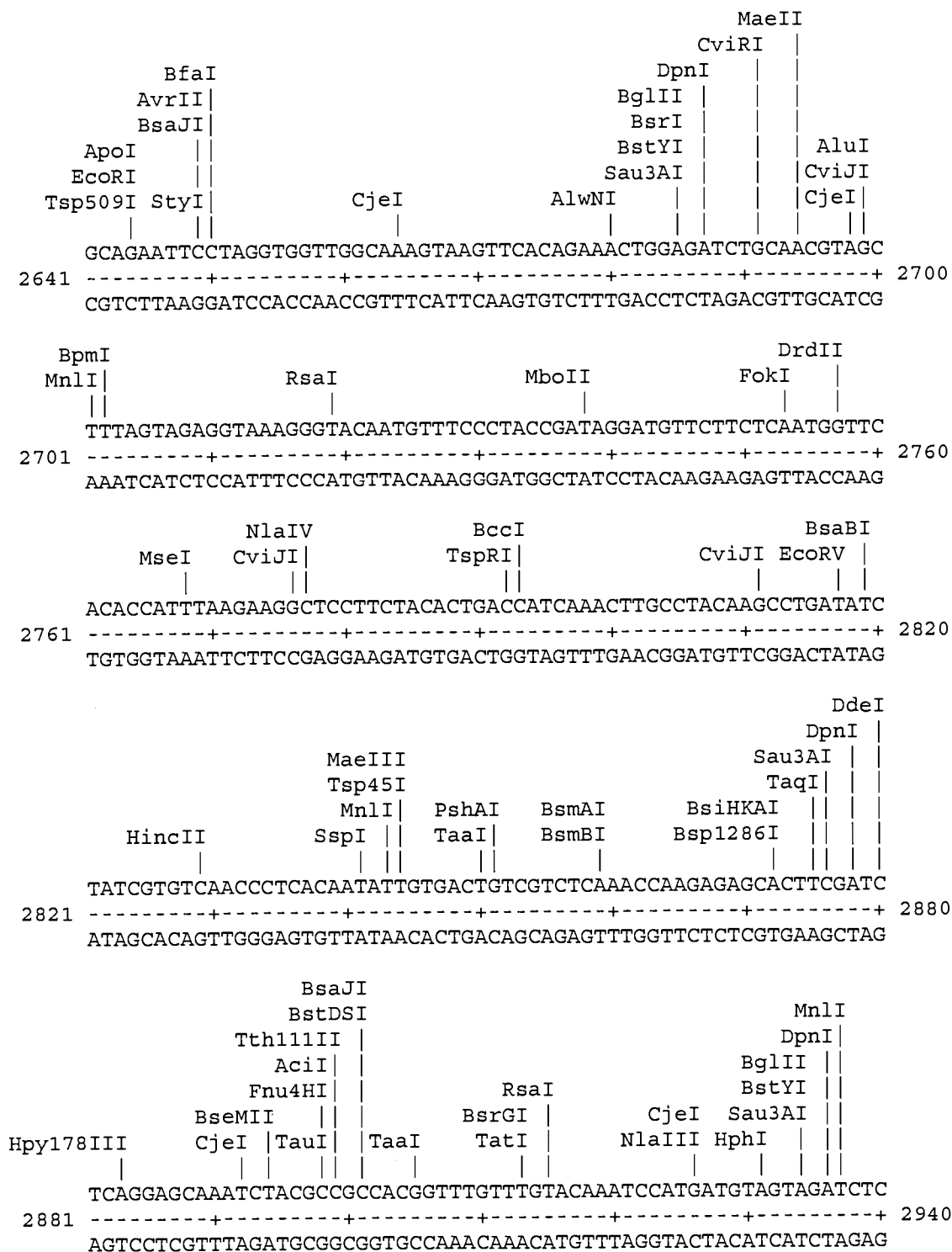


Fig. 26 (con't)

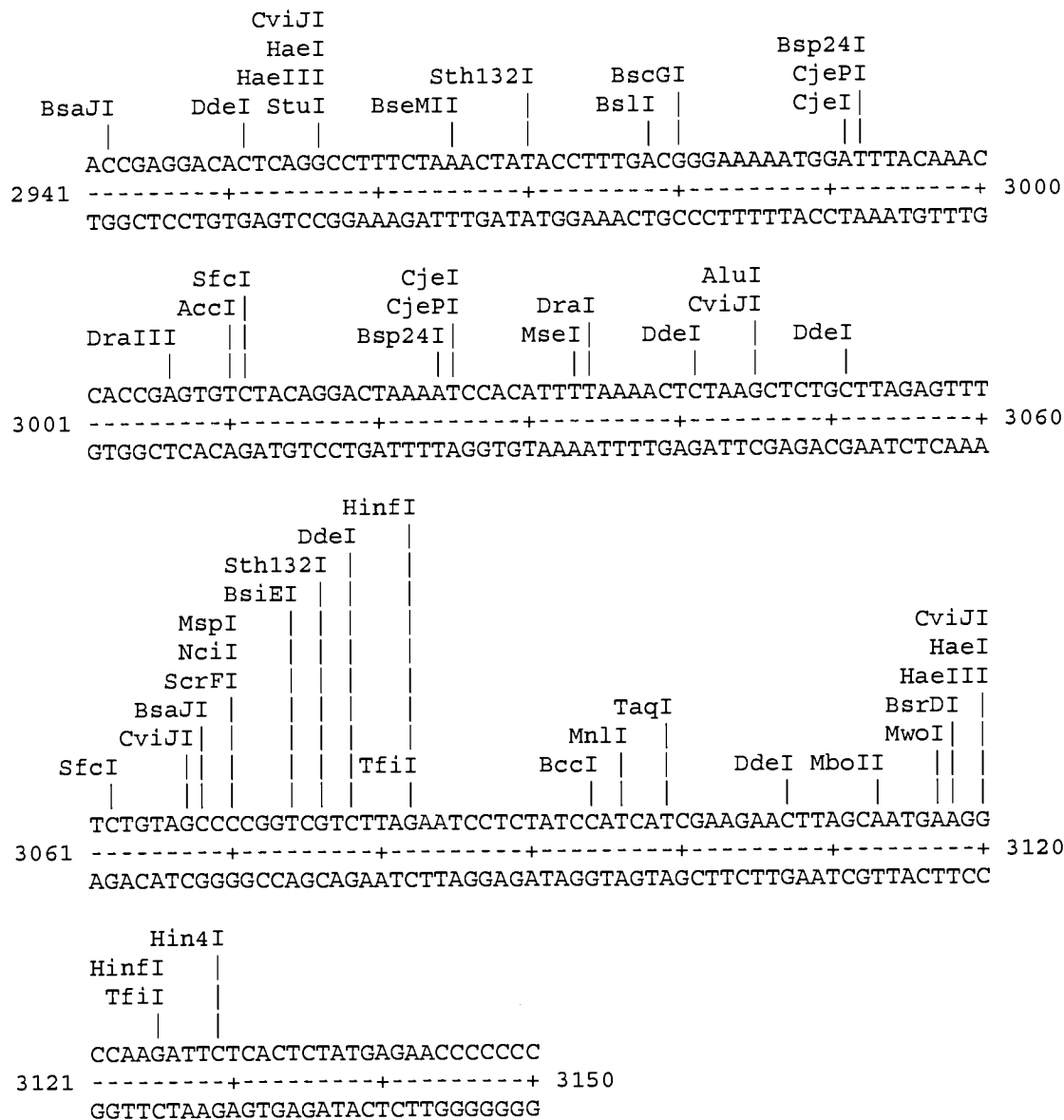


Figure 27: CPN100397

```

1  MKIPLRFLLI  SLVPTLSMSN  LLGAATTEEL  SASNSFDGTT  STTSFSSKTS
51 SATDGTNYVF  KDSVVIENVP  KTGETQSTSC  FKNDAAAGDL  NFLGGGFSFT
101 FSNIDATTAS  GAAIGSEAAAN  KTVTLSGFSA  LSFLKSPAST  VTNGLGAINV
151 KGNLSLLDND  KVLIQDNFST  GDGGAINCAG  SLKIANNKSL  SFIGNSSSTR
201 GGAIHTKNLT  LSSGGETLFQ  GNTAPTAAGK  GGAIAIADSG  TLSISGDSGD
251 IIFEGNTIGA  TGTVSHSAID  LGTSAKITAL  RAAQGHTIYF  YDPITVTGST
301 SVADALNINS  PDTGDNKEYT  GTIVFSGEKL  TEAEAKDEKN  RTSKLLQNVA
351 FKNGTVV LKG  DVVLSANGFS  QDANSKLIMD  LGTSLVANTE  SIELTNLEIN
401 IDSLRNGKKI  KLSAATAQKD  IRIDRPVVLA  ISDESFYQNG  FLNEDHSYDG
451 ILELDAGKDI  VISADSRSID  AVQSPYGYQG  KWTINWSTDD  KKATVSWAKQ
501 SFNPTAEQEA  PLVPNLLWGS  FIDVRSFQNF  IELGTEGAPY  EKRFVWAGIS
551 NVLHRSGREN  QRKFRHVSGG  AVVGASTRMP  GGDTLSLGFA  QLFARDKDYF
601 MNTNFAKTYA  GSLRLQHDAS  LYSVVSILLG  EGGLREILLP  YVSKTLPCSF
651 YGQLSYGHTD  HRMKTESLPP  PPPTLSTDHT  SWGGYVWAGE  LGTRVAVENT
701 SGRGFFQEYT  PFVKVQAVYA  RQDSFVELGA  ISRDFSDSL  YNLAIPLGK
751 LEKRFAEQYY  HVVAMYSPOV  CRSNPKCTTT  LLSNQGSWKT  KGSNLARQAG
801 IVQASGFRSL  GAAAELEFGN  GFEWRGSSRS  YNVDAGSKIK  F

```

Possible T cell epitope:

516 LLWGSFIDV

Possible B cell epitope:

554 HRSGRENQRKFRHV

Figure 28: CPN100421

1 MPPLNADDVL PRDHLSGDSF SDTYPDITTQ AIILIFLALS PFLVMLLTSY  
51 LKIIITLVLL RNALGVQOTP PSQVLNGIAL ILSIYVMFPT GVAMYKDARK  
101 EIEANTIPQS LFTAEGAETV FVALNKSKEP LRSFLIRNTP KAQIQSFYKI  
151 SQKTFPSEIR AHLTASDFVI IIPAFIMGQI KNAFEIGVLI YLPFFVIDLV  
201 TANVLVAMQM MMLSPLSISL PLKLLLIVMV DGWTLLLOGL MISFK

Possible T cell epitope:

188 VLIYLPFFV

Possible B cell epitope:

125 NKSKEPLR

Figure 29: CPN100422

1 MKFFSLIFKD DDVSPNKKVL SPEAFSAFLD AKELLEKTKA DSEAYVAETE  
51 QKCAQIRQEA KDQGFKEGSE SWSKQIAFLE EETKNLRIRV REALVPLAIA  
101 SVRKIIIGKEL ELHPETIVSI ISQALKELTQ NKHIIISVNP KDLPLVEKSR  
151 PELKNIVEYA DSLILTAKPD VTPGGCIIET EAGIINAQLD VQLDALEKAF  
201 STILKAKNPV DEPSETSSST DSSSLSDNDQD KKE

Possible T cell epitope:

163 LILTAKPDV

Possible B cell epitope:

226 SNDQDKKE



Figure 30: CPN100424

```
1  MTLCCCTSCN SRSLIVHGLP GREANEIVVL LVSKGVAAQK LPQAAAATAG
51 AATEQMWDIA VPSAQITEAL AILNQAGLPR MKGTSLLDLF AKQGLVPSEL
101 QEKIRYQEGE SEQMASTIRK MDGVVDASVQ ISFTTENEDN LPLTASVYIK
151 HRGVLDNPNS IMVSKIKRLI ASAVPGLVPE NVSVVSDRAA YSDITINGPW
201 GLTEEIDYVS VWGIILAKSS LTKFRLIFYV LILILFVISC GLLWVIWKTH
251 TLIMTMGGTK GFFNPTPYTK NALEAKKAEG AAADKEKKED ADSQGESKNA
301 ETSKDKSSDK DAPEGSNEIE GA
```

Possible T cell epitope:

201 GLTEEIDYV

Possible B cell epitope:

284 DKEKKEDADSQGESKNAETSKDKSSDKDAPEGSNEIE

Figure 31: CPN100426

1 MTIRVRNLAY SVNKKKILDG VTFSLERGHI TLFVGKSGSG KTMILRALAG  
51 LVQPTQGDIW IEGEAPALVF QQPELFSHMT VLGNCNTHPQI HIKGRSTEEA  
101 REKAFELLHL LDIEEVAKNY PDQLSGGQKQ RVAIVRSLCM DKHTLLFDEP  
151 TSALDPFATA SFRHLLETLR DQELTVGLTT HDMQFVHSCL DRIYLIDQGT  
201 VAGVYDKRDG ELDSGHPLSK YIHSAQ

Possible T cell epitope:

145 LLFDEPTSA

Possible B cell epitope:

205 YDKRDGE

Figure 32: CPN100508

```
1 MKRPFFTYLC IIFYGSCASL SLHAGLSFPE VRGATAAVVH ADGKVFYDK
51 DIDAVIYPAS MTKIATALFI LKHYPTVLDL LIKVKQDAIA SITPQAKKQS
101 GYRSPPHWLE TDGSTIQLHL REELLGWDLF HALLVCSAND AANVLAMACC
151 GSVEKFMDKL NFFLKEEIGC THTHFNPNHG LHHPNHYTTT RDLISIMRCA
201 LKEPPFRGVI STTSYKIGAT NLHGERILSP TNKLLLPGST YHYPPALGGK
251 TGTTKTAGKN LIMAAEKNNR LLVTIATGYS GPVSDLYQDV IALCETVFNE
301 PLLRKELVPP SDCLQLEIAN LGKLSCPLPE GLYYDFYASE DREPLSVSFI
351 AHADAFPIEQ GDLLGHWFY DDEGKKISSQ PFYAPCRFER TIKPWKLYMK
401 RVFTSYRTYM SITMLLMYFR IRKHRKYKNL KHYSKI
```

Possible T cell epitope:

156 FMDKLNFFL

Possible B cell epitope:

422 RKHRKYKN

Figure 33: CPN100515

```

  1 MASNPIQLQIE DLSITLAKQR QQYPIVQSL S FTINEGQTLA IIGESGSGKS
 51 VSAHAILRLL PCPPFSVSGQ VNFQGHNLLT ASRSIQKKII GTEISMIFQN
101 PQASLNPVFT IEQQFREIIH THLALTA EVA KEKMLYALEE TGFHDPRLCL
151 NLYPHQLSGG MLQRICIAM A LLCSPKLLIA DEPTALDVS VQYQILQLLK
201 TLQKKTGMSL LIITHNMGVV AETADDVLVL YAGRMVECAP AVQMFHNPSH
251 PYTRDLLASR PSLQPQQLGS FNPIPGQPPH YTAFPSCCRY HPRCSKILNR
301 CSAEAEPIYP VREGHKVRVG CMTTNFPQPL IQATSLTKHY YKRSFWFQ GK
351 TIASRPVDDV SFSLYSRRV GLIGESGSGK STLALALAGL LPLTSGFLT F
401 NGTPIKLH SK HGRHQLRSQV RLVFQNPQAS LNPRKTILDS LGHSLLYHKL
451 VPKEKVLATV REYLELVGLS EEFYRYPHQ LSGGQQQQRVS IARALLGV PQ
501 LIICDEIVSA LDLSIAQIL NMLAELQKKL SLTYLFISHD LAVVRSFCTE
551 VFIMYKGQIV EKGNTKRIFS DPQHPYTRML LNAQLPETPD QRQSKPIFQE
601 YHKDSEESCS TGCYFYNRCP QKQEACKSEI IPNQGDHHT YRCIH

```

Possible T cell epitope:

59      LLPCPPFSV

Possible B cell epitopes:

18      KQRQQY  
 587     ETPDQRQSK

Figure 34: CPN100538

```
1  MPGIEKAATT VAVPQDKSEE EKVKERLTRK ELTCEDLKDN GYTVNFEDIS
51 ILELLQFVSK ISGTNFVFDN NDLQFNVTIV SHDPTSVDDL STILLQVLKM
101 HDLKVVEQGN NVLIYRNPHL SKLSTVVTDS SLKETCEAVV VTRVFRLYRR
151 QPSAAVNIIQ PLLSHDAIVS ASEATRHVII SDIAGNVDKV SDLLAALDCP
201 GTSVDMTEYE VKYANPAALV SYCQDVLGTL AEDDAFQPMFI QPGTNKIFVV
251 SSPRLANKAE QLLKSLDVPE MAHTLDDPAS TALALGGTGT TSPKSLRFFM
301 YKLKYQNGEV IANALQDIGY NLYVTTAMDE DFINTLNSIQ WLEVNSIVI
351 IGNQGNVDRV IGLLNGLDLP PKQVYIEVLI LDTSLEKSWD FGVQWVALGD
401 EQSKVAYASG LLNNTGIATP TKATVPPGTP NPGSIPLPTP GQLTGFSDML
451 NSSSAFGLGI IGNVLSHK GK SFLT LGLLS ALDQDGD TVI VLNPRIMAQD
501 TQQASFFVGQ TVPYQTIKYY IQETGTVTQN IDYEDIGVNL VVTSTVAPNN
551 VVTLQIEQTI SELHSASGSL TPVTDKTYAA TRLQIPDGCF LVMSGHIRDK
601 TTKVVS GVPL LNSIPLIRGL FSRTIDQRQK RNIMMFIKPK VISSFEEGTR
651 VTNKEGYRYN WEADEGSMQV APRHAPECQG PPSLQAESDF KIIEIEAQ
```

Possible T cell epitope:

50     SILELLQFV

Possible B cell epitopes:

15     QDKSEEEK

626    DQRQKRN

Figure 35: CPN100557

```
1  MSRKDNEVSL ARSIFNILSG TFCSRITGIF REIAMATYFG ADPIVAAFVL
51 GFRTVFFLRK ILGGLILEQA FIPHFEFLRA QSLDRAAFFR RRFSLIKGS
101 TIIFTLLIEA VLWVFFNNVE EGTYDMILLT MILLPCGIFL MMYNVNGALL
151 HCGNKFFGVG LAPVVVNIW IFFVIAARHS DPRERIIGLS VALVIGFFFE
201 WLITVPGVWK FLLEAKSPPQ EHDSVRALLA PLSLGILTSS IFQLNLLSDI
251 CLARYVHEIG PLYLMYSLKI YQLPIHLFGF GVFTVLLPAI SRCVQREDHE
301 RGLKLMKFVL TLTMSVMIIM TAGLLLLLALP GVRVLYEHGL FPQSAVYAIV
351 RVLRGYGASI IPMALAPLVS VLFYAQRQYA VPLFIGIGTA LANIVLSLVL
401 GRWVLKDVSG ISYATSITAW VQLYFLWYYS SKRLPMYSKL LWESIRRSIK
451 VMGTTMLACM ITLGLNILTQ TTYVIFLNPL TPLAWPLSSI TAQAIAFLSE
501 SCIFLAFLFG FAKLLRVEDL INLASFEYWR GQRGLLQRQH VMQDTQN
```

Possible T cell epitope:

111 VLWVFFNNV

Possible B cell epitopes:

1 MSRKDNE  
295 QREDHERG

Figure 36: CPN100622

```

1 MKTSRNKQCK ITDPLSKSSF FVGALILGKT TILLNATPLS DYFDNQANQL
51 TTLFPLIDTL TNMTPYSHRA TLFGVRDDTN QDIVLDHQNS IESWFENFSQ
101 DGGALSKSL AITNTKNQIL FLNSFAIKRA GAMVVDGNFD LSENHGSIIF
151 SGNLSFPNAS NFADTCTGGA VLCSKNVTIS KNQGTAYFIN NKAKSSGGAI
201 QAAIINIKDN TGPCLFFNNA AGGTAGGALF ANACRIENNS QPIYFLNNQS
251 GLGGAIRVHQ ECILTKNTGS VIFNNNFAME ADISANHSSG GAIYCISCSI
301 KDNPGIAAFD NNTAARDGGA ICTQSLTIQD SGPVYFTNNQ GTWGGAIMLR
351 QDGACTLFAD QGDIIFYNNR HFKDTFSNHV SVNCTRNVS LTVGASQGHSA
401 TFYDPILQRY TIQNSIQKFN PNPEHLGTIL FSSTYIPDTS TSRDDFISHF
451 RNHIGLYNGT LALEDRAEWK VYKFDQFGGT LRLGSRVFS TTDDEEQSSSS
501 VGSVININNL AINLPSILGN RVAPKLWIRP TGSSAPYSED NNPIINLSGP
551 LSLDDENLD PYDTADLAQP IAEVPLLYLL DVTAKHINTD NFYPEGLNTT
601 QHYGYQGVWS PYWIETITTS DTSS EDTVNT LHRQLYGDWT PTGYKVN PEN
651 KGDIALSAFW QSFHNL FATL RYQTQQGQIA PTASGEATRL FVHQNSN NDA
701 KGFHMEATGY SLGTTSNTAS NHSFGVNFSQ LFSNLYESHS DNSVASHTTT
751 VALQINN PWL QERFSTSASL AYSYSNHHIK ASGYSGKIQT EGKCYSTTLG
801 AALSCSLSLQ WRSRPLHFTP FIQAI AVRSN QTAFQESGDK ARKFSVHKPL
851 YNLTVPLGIQ SAWESKFRLP TYWNIELAYQ PVLYQQNPEI NVSLESSGSS
901 WLLSGTTLAR NAI AFKGRNQ IFIFPKLSVF LDYQGSVSSS TTTHYLHAGT
951 TFKF

```

Possible T cell epitope:

119 ILFLNSFAI

Possible B cell epitopes:

2 KTSRNKQ  
647 NPENKG  
694 QNSNND AK

Figure 37: CPN100626

```

1  MQVFPKVTLS LDYSADISSS TLSHYLNVAS RMRFLTISDQ NRKIKEPLVS
51 KTPPKFLFYI GNFTACMFGM TPAVYSLQTD SLEKFALERD EEFRTSFPLL
101 DSLSTLTGFS PITTFVGNRH NSSQDIVLSN YKSIDNILL WTSAGGAVSC
151 NNFLLSNVED HAFFSKNLAI GTGGAIACQG ACTITKNRGP LIFFSNRGLN
201 NASTGGETRG GAIACNGDFT ISQNGTFFYF VNNSVNNWGG ALSTNGHCRI
251 QSNRAPLLFF NNTAPSGGGA LRSENTTISD NTRPIYFKNN CGNNGGAIQT
301 SVTVAIKNNS GSVIFNNNTA LSGSINSGNG SGGAIYTTNL SIDDNPGTIL
351 FNNNYCIRDG GAICTQFLTI KNSGHVYFTN NOGNWGGALM LLQDSTCLLF
401 AEQGNIAFQN NEVFLTTFGR YNAIHCTPNS NLQLGANKGY TTAFFDPIEH
451 QHPTTNPLIF NPNANHOGTI LFSSAYIPEA SDYENNFISS SKNTSEL RNG
501 VLSIEDRAGW QFYKFTQKGG ILKLGHAA SI ATTANSETPS TSVGSQVIIN
551 NLAINLPSIL AKGKAPTLWI RPLQSSAPFT EDNNPTITLS GPLTLLNEEN
601 RDPYDSIDLS EPLQNIHLLS LSDVTARHIN TDNFHPESLN ATEHYGYQGI
651 WSPYWVETIT TTNNASIETA NTLYRALYAN WTPLGYKVNP EYQGD LATTP
701 LWQSFHTMFS LLRSYNRTGD SDIERPFLEI QGIADGLFVH QNSIPGAPGF
751 RIQSTGYSLQ ASSETSLHOK ISLGFAQFFT RTKEIGSSNN VSAHNTVSSL
801 YVELPWFQEA FATSHSLAYG YGDHHLHAYI RHIKNRAEGT CYSHTLAAAI
851 GCSFPWQQKS YLHLSPFVQA IAIRSHQTAF EEIGDNPRKF VSQKPFYNLT
901 LPLGIQGWQ SKFHVPTWT LELSYQPVLY QQNPQIGVTL LASGGSWDIL
951 GHNYVRNALG YKVHNQTALF RSLDLFLDYQ GSVSSSTSTH HLQAGSTLKF

```

Possible T cell epitope:

56 FLFYLGNFT

Possible B cell epitopes:

39 DQNRKIK

597 NEENRDPYD



Figure 38: CPN100628

```
1  MLLPFTFVLA NEGLQLPLET YITLSPEYQA APQVGFTTHQ NQDLAIVGNH
51 NDFILDYKYY RSNGGALTCK NLLISENIGN VFFEKNVCPN SGGAIYAAQN
101 CTISKQNQNYA FTTNLVSDNP TATAGSLG G ALFAINC SIT NNLGQGT FVD
151 NLALNKG GAL YTETNLSIKD NKGPIIIKQN RALNSDSLGG GIYSGNSLNI
201 EGN SGAIQIT SNSSGSGGGI FSTQTLTISS NKKLIEISEN SAFANNYGSN
251 FNP GGGGLTT TFCTILNNRE GVLFNNNQSQ SNGGAIHAKS I I I KENGPVY
301 FLNNTATRGG ALLNLSAGSG NGSFILSADN GDII FNNNTA SKHALNPPYR
351 NAIHSTPNMN LQIGARPGYR VLFYDPIEHE LPSSFPILFN FETGHTGTVL
401 FSGEHVHQNF TDEMNFYSYL RNTSELRQGV LAVEDGAGLA CYKFFQGGT
451 LLLGQGAVIT TAGTIPTPSS TPTTVGSTIT LN HIAIDLPS ILSFQAQAPK
501 IWIYPTKTGS TYTEDSNPTI TISGTLTLRN SNNEDPYDSL DLSHSLEKVP
551 LLYIVDVAAQ KINSSQLDLS TLNSGEHYGY QGIWSTYWVE TTTITNPTSL
601 LGANTKHKLL YANWSPLGYR PHPERRGEFI TNALWQSAYT ALAGLHSLSS
651 WDEEKGHAAS LQIGLLVHQ KDKNGFKGFR SHMTGYSATT EATSSQSPNF
701 SLGFAQFFSK AKEHESQNST SSHHYFSGMC IAKYSLQ RVI RLSVSLAYMF
751 TSEHTHTMYQ GLLEGNSQGS FHNHTLAGAL SCVFLPQPHG ESLQIYPFIT
801 ALAIRGNLAA FQESGDHARE FSLHRPLTDV SLPVGIRASW KNHHRVPLVW
851 L TEISYRSTL YRQDPELH SK LLISQGTWTT QATPVTYNAL GIKVKNTMQV
901 FPKVTLSLDY SADISSSTLS HYLNVASRMR F
```

Possible T cell epitope:

1 MLLPFTFVL

Possible B cell epitopes:

```
38  HNQNQ
619 YRPHPERRG
669 HQDKNG
```

Figure 39: CPN100630

```

1  MPLSFKSSSF  CLLACLCSAS  CAFAETRLGG  NFVPPITNQG  EEILLTSDFV
51  CSNFLGASFS  SSFINSSSNL  SLLGKGLSLT  FTSCQAPTNS  NYALLSAAET
101 LTFKNFSSIN  FTGNQSTGLG  GLIYGKDIVF  QSIKDLIFTT  NRVAYSPASV
151 TTSATPAITT  VTTGASALQP  TDSLTVENIS  QSIKFFGNLA  NFGSAISSSP
201 TAVVKFINNT  ATMSFSHNFT  SSGGGVIYGG  SSLLFENNSG  CIIFTANSCV
251 NSLKGVTSS  GTYALGSGGA  ICIPTGTFEL  KNNQKGCTFS  YNGTPNDAGA
301 IYAETCNIVG  NQGALLLDSN  TAARNNGAIC  AKVLNIQGRG  PIEFSRNRAE
351 KGGAIFIGPS  VGDPKQTST  LTIASEGDI  AFQGNMLNTK  PGIRNAITVE
401 AGGEIVSLSA  QGGSRLVFD  PITHSLPTTS  PSNKDITINA  NGASGSVVFT
451 SKGLSSTELL  LPANTTTILL  GTVKIASGEL  KITDNAVVNV  AGFATQSGSQ
501 LTLGSGGTLG  LATPTGAPAA  VDFTIGKLAF  DPFSFLKRDF  VSASVNAGTK
551 NVTLTGALVL  DEHDVTDLYD  MVSLQSPVAI  PIAVFKGATV  TKTGFPDGEI
601 ATPSHYGYQG  KWSYTWSRPL  LIPAPDGGFP  GGPSPSANTL  YAVWNSDTLV
651 RSTYILDPER  YGEIVSNSLW  ISFLGNQAFS  DILQDVLLID  HPGLSITAKA
701 LGAYVEHTPR  QGHEGFSGRY  GGYQAALSMN  YTDHTTLGLS  FGQLYGKTNA
751 NPYDRCSEQ  MYLLSFFGQF  PIVTQKSEAL  ISWKAAYGYS  KNHLNTTYLR
801 PDKAPKSQGG  WHNNSYYVLI  SAEHPFLNWC  LLTRPLAQAW  DLSGFISAEF
851 LGGWQSKFTE  TGDQRSFSR  GKGYNVSLPI  GCSSQWFTPF  KKAPSTLTIK
901 LAYKPDYRV  NPHNIVTVVS  NQESTSISGA  NLRRHGLFVQ  IHDVVDLTED
951 TQAFNLNYTFD  GKNGFTNHRV  STGLKSTF

```

Possible T cell epitope:

936 GLFVQIHDV

Possible B cell epitopes:

281 KNNQGK  
 345 SRNRAEK  
 707 HTPRQGHE